

Sat Jun ' 8 10:13:30 2002

us-09-729-674-1.rng

SEA ID NO: 1  
Database: N.GeneSeq-032802  
AC NO: AAX60801

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 7, 2002, 21:01:30 ; Search time 562.42 Seconds  
(without alignments)  
11817.104 Million cell updates/sec

Title: US-09-729-674-1

Perfect score: 3871

Sequence: 1 ttctctctccctccctt.....ataaaaaaaaaaaaaa 3871

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

N.GeneSeq-032802: \*  
1: /SIDSI/gcgdata/geneSeq/geneSeqn-emb1/NA1980.DAT: \*  
2: /SIDSI/gcgdata/geneSeq/geneSeqn-emb1/NA1981.DAT: \*  
3: /SIDSI/gcgdata/geneSeq/geneSeqn-emb1/NA1982.DAT: \*  
4: /SIDSI/gcgdata/geneSeq/geneSeqn-emb1/NA1983.DAT: \*  
5: /SIDSI/gcgdata/geneSeq/geneSeqn-emb1/NA1984.DAT: \*  
6: /SIDSI/gcgdata/geneSeq/geneSeqn-emb1/NA1985.DAT: \*  
7: /SIDSI/gcgdata/geneSeq/geneSeqn-emb1/NA1986.DAT: \*  
8: /SIDSI/gcgdata/geneSeq/geneSeqn-emb1/NA1987.DAT: \*  
9: /SIDSI/gcgdata/geneSeq/geneSeqn-emb1/NA1988.DAT: \*  
10: /SIDSI/gcgdata/geneSeq/geneSeqn-emb1/NA1989.DAT: \*  
11: /SIDSI/gcgdata/geneSeq/geneSeqn-emb1/NA1990.DAT: \*  
12: /SIDSI/gcgdata/geneSeq/geneSeqn-emb1/NA1991.DAT: \*  
13: /SIDSI/gcgdata/geneSeq/geneSeqn-emb1/NA1992.DAT: \*  
14: /SIDSI/gcgdata/geneSeq/geneSeqn-emb1/NA1993.DAT: \*  
15: /SIDSI/gcgdata/geneSeq/geneSeqn-emb1/NA1994.DAT: \*  
16: /SIDSI/gcgdata/geneSeq/geneSeqn-emb1/NA1995.DAT: \*  
17: /SIDSI/gcgdata/geneSeq/geneSeqn-emb1/NA1996.DAT: \*  
18: /SIDSI/gcgdata/geneSeq/geneSeqn-emb1/NA1997.DAT: \*  
19: /SIDSI/gcgdata/geneSeq/geneSeqn-emb1/NA1998.DAT: \*  
20: /SIDSI/gcgdata/geneSeq/geneSeqn-emb1/NA1999.DAT: \*  
21: /SIDSI/gcgdata/geneSeq/geneSeqn-emb1/NA2000.DAT: \*  
22: /SIDSI/gcgdata/geneSeq/geneSeqn-emb1/NA2001A.DAT: \*  
23: /SIDSI/gcgdata/geneSeq/geneSeqn-emb1/NA2001B.DAT: \*  
24: /SIDSI/gcgdata/geneSeq/geneSeqn-emb1/NA2002.DAT: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the local score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3870.2	100.0	3871	20	AAX60801 Human secreted pro
2	3870.2	100.0	3871	22	AAS59207 Human CDNA encodin
3	3870.2	100.0	3871	24	ABA90876 Human polynucleoti
4	2051.2	53.0	2549	21	AAAF16225 Human prostate can
5	1955.4	50.5	1961	22	AAH14525 Human CDNA sequenc
6	1474.4	38.1	1586	21	AAH31138 Human CDNA clone (
7	695.6	18.0	818	22	AAH06808 Human colon cancer
8	466.4	12.0	618	21	AAZ80131 Human colon cancer
9	406.4	10.5	408	21	AAH30585 Human colon cancer

10	371.6	9.6	401	22	AAH50786 Human tumour assoc
11	369.4	9.5	400	16	AAT21281 Human gene signatu
12	355.6	9.2	407	22	AAH50808 Human tumour assoc
13	340	8.8	351	22	AAH11840 Human CDNA clone (
14	337.2	8.7	373	21	AAH43677 Human secreted exp
15	258.6	6.7	279	23	AAS58007 Human CDNA #683 encodin
16	239.2	6.2	6670	22	AAH02846 Human reproductive
17	239.2	6.2	7713	22	AAH02845 Human reproductive
18	238.4	6.2	8658	22	AAH69231 Human immune/haema
19	238	6.1	295	22	AAH82995 Human immune/haema
20	235.4	6.1	32351	21	AAH21307 Human low adenosin
21	235.4	6.1	32351	21	AAH35185 Human adenosine re
22	235.4	6.1	40298	21	AAH21311 Human low adenosin
23	235.4	6.1	40298	21	AAH35189 Human adenosine re
24	234.6	6.1	4275	22	AAH19408 Human nervous syst
25	234.6	6.1	4883	22	AAH79410 Human immune/haema
26	234.6	6.1	16997	22	AAH80033 Human immune/haema
27	234.6	6.1	16997	22	AAH86146 Human immune/haema
28	233.8	6.0	9620	22	AAH06207 Human liver cell s
29	233.4	6.0	6138	22	AAH57456 Human CDNA clone (
30	233.2	6.0	763	22	AAH06411 Human CDNA sequenc
31	232.6	6.0	2283	22	AAH14438 Human immune/haema
32	232.6	6.0	5979	22	AAH78985 Human immune/haema
33	232.6	6.0	5979	22	AAH80951 Human immune/haema
34	232.2	6.0	532	22	AAH71264 Human cervical can
35	232.2	6.0	567	22	AAH69839 Human cervical can
36	232	6.0	3775	22	AAH61061 Human polynucleoti
37	231.4	6.0	1449	22	AAH86057 Human immune/haema
38	231.4	6.0	1450	22	AAH86062 Human immune/haema
39	230.8	6.0	2847	22	AAH18224 Human nervous syst
40	230.6	6.0	2847	22	AAH82073 Human immune/haema
41	230.6	6.0	2108	22	AAH76677 Human immune/haema
42	230.6	6.0	2108	22	AAH76678 Human immune/haema
43	230.2	5.9	646	22	AAH7915 Human immune/haema
44	230.2	5.9	1146	22	AAH90464 Human digestive sy
45	230	5.9	13559	22	AAH15144 Human nervous syst

ALIGNMENTS

RESULT 1	
AAX60801	
ID AAX60801 standard; DNA: 3871 BP.	
XX	
AC AAX60801:	
XX	
DT 09-AUG-1999 (first entry)	
XX	
DE Human secreted protein encoding DNA (clone bd306-7).	
XX	
KW Secreted protein; kidney; lung; brain; blood; testis; bone marrow;	
KW nutritional activity; cytokine; cell proliferation; immune stimulation;	
KW hematopoiesis regulation; tissue growth; thrombolytic; gene therapy;	
KW anti-inflammatory; tumour invasion; ss.	
XX	
OS Homo sapiens.	
XX	
PN WO926961-A1.	
XX	
PD 03-JUN-1999.	
XX	
PF 24-NOV-1998; 98WO-US25149.	
XX	
PR 23-NOV-1998; 98US-0197886.	
XX	
PR 26-NOV-1997; 97US-0066804.	
XX	
PA (GEMV ) GENETICS INST INC.	
XX	
PI Agostino M, Clark HF, Collins-Racie LA, Evans C;	
PI Fechtel K, Jacobs K, Lavallee ER, McCoy JM, Werberg D;	
PI Steinhilber RJ, Treacy M, Wong GG;	
XX	



|||||  
Db 1681 ccaagcgtgctcgaactcttgacttgaactgaatgacatcctgcttggcctccccaagtgc 1740  
OY 1741 tgggattacagggctgagccacatgcgcggccctcttcacacttaagacctctctc 1800  
Db 1741 tgggattacagggctgagccacatgcgcggccctcttcacacttaagacctctctc 1800  
OY 1801 tatctacatctgtttctcacactcatccctgtctctctcatgttcaactgtcttc 1860  
Db 1801 tatctacatctgtttctcacactcatccctgtctctctctcatgttcaactgtcttc 1860  
OY 1861 ccaatgttcaatgctgccttcttcaacttggttgaaaggcgagctctctgtgctg 1920  
Db 1861 ccaatgttcaatgctgccttcttcaacttggttgaaaggcgagctctctgtgctg 1920  
OY 1921 ttttttggttttcccgaaataatcagatatttttaataagaanaaatctctagaa 1960  
Db 1921 ttttttggttttcccgaaataatcagatatttttaataagaanaaatctctagaa 1960  
OY 1981 gatgataattgtgaaacccctctgtgcttattgtcttccagatttagtctcttc 2040  
Db 1981 gatgataattgtgaaacccctctgtgcttattgtcttccagatttagtctcttc 2040  
OY 2041 tcccatcgcgggaaagaatggctggaagacataggtctaaattctccagcctcaaaatgctc 2100  
Db 2041 tcccatcgcgggaaagaatggctggaagacataggtctaaattctccagcctcaaaatgctc 2100  
OY 2101 ttaactgtgtctgaactgttcccaatttgaaccccatgnaaaaagaattgagtagagag 2160  
Db 2101 ttaactgtgtctgaactgttcccaatttgaaccccatgnaaaaagaattgagtagagag 2160  
OY 2161 tgtagagtgagaaatgttgctctttgcccactttgtcatctccaaaatlaaagaagtgcg 2220  
Db 2161 tgtagagtgagaaatgttgctctttgcccactttgtcatctccaaaatlaaagaagtgcg 2220  
OY 2221 ccgattcccatlttgaggaacaaatgtcttgaatlaaagtlccgagtltygnaaaggaaagac 2280  
Db 2221 ccgattcccatlttgaggaacaaatgtcttgaatlaaagtlccgagtltygnaaaggaaagac 2280  
OY 2281 caaagcgtctagttattcatcatcttcttcagtaataataattagagtaactactgtctgc 2340  
Db 2281 caaagcgtctagttattcatcatcttcttcagtaataataattagagtaactactgtctgc 2340  
OY 2341 taagcatctgaccctgggaactagagatatcttcacagaataacaggaagttccctgtgct 2400  
Db 2341 taagcatctgaccctgggaactagagatatcttcacagaataacaggaagttccctgtgct 2400  
OY 2401 catggaacttacaattctacagagggaagaagatagccaatacatatagaatataatatac 2460  
Db 2401 catggaacttacaattctacagagggaagaagatagccaatacatatagaatataatatac 2460  
OY 2461 aaggtatcatgtatgtataatctgctgtggaaaaaataaagcagggagggagtaagaa 2520  
Db 2461 aaggtatcatgtatgtataatctgctgtggaaaaaataaagcagggagggagtaagaa 2520  
OY 2521 tctctgagatgagagccctgaatttaaaatgggctcaactggaatggaactgtgaagaa 2580  
Db 2521 tctctgagatgagagccctgaatttaaaatgggctcaactggaatggaactgtgaagaa 2580  
OY 2581 gacgtttagggaaatggaatcccttgacaagagcatctccaggcagaggaacaaagtgtgacatg 2640  
Db 2581 gacgtttagggaaatggaatcccttgacaagagcatctccaggcagaggaacaaagtgtgacatg 2640  
OY 2641 ccccaagaatgagaactgtgctctacgtggtcaggaagaagacagaggaaccaagagatgcg 2700  
Db 2641 ccccaagaatgagaactgtgctctacgtggtcaggaagaagacagaggaaccaagagatgcg 2700  
OY 2701 tgggcaaggggtagaatggaagggagggcggctggggaggaacaggtgtgtgaaggccttgg 2760  
Db 2701 tgggcaaggggtagaatggaagggagggcggctggggaggaacaggtgtgtgaaggccttgg 2760  
OY 2761 ctctctgaagtgaatggaagcaactggaagggttctgaacagagaggtgcttgatgat 2820  
|||||

Db 2761 ctctctgaagtgaatggaagcaactggaagggttctgaacagagaggtgccttgatgat 2820  
OY 2821 ttatatcttgaaaggtcatctctagctgcgaataatgttgaaanaactttagtgaacaaggc 2880  
Db 2821 ttatatcttgaaaggtcatctctagctgcgaataatgttgaaanaactttagtgaacaaggc 2880  
OY 2881 agaaaggaagggagaagactgtttagaagctactgcaaggttccaggtcttgagcctggc 2940  
Db 2881 agaaaggaagggagaagactgtttagaagctactgcaaggttccaggtcttgagcctggc 2940  
OY 2941 cacaagcaacagcagtggtcaaatatctagatttatttgaaagaagccaatagatgttc 3000  
Db 2941 cacaagcaacagcagtggtcaaatatctagatttatttgaaagaagccaatagatgttc 3000  
OY 3001 tgaagatttgaatgttgagtgctgaagaaaggaagtttaagatgacatlaaagtttttg 3060  
Db 3001 tgaagatttgaatgttgagtgctgaagaaaggaagtttaagatgacatlaaagtttttg 3060  
OY 3061 gctctgaaatagcaagaaagatggaatlaacagttactgaataatagggaaatggtctggt 3120  
Db 3061 gctctgaaatagcaagaaagatggaatlaacagttactgaataatagggaaatggtctggt 3120  
OY 3121 aagtaaggaaattgtgtcacaagcagctgtctgtgttgaaatvggaaggtctctgcgca 3180  
Db 3121 aagtaaggaaattgtgtcacaagcagctgtctgtgttgaaatvggaaggtctctgcgca 3180  
OY 3181 aatcaagtgagaatctctcagaagctcaggtctgaagcaagactcgaagaacagggatcga 3240  
Db 3181 aatcaagtgagaatctctcagaagctcaggtctgaagcaagactcgaagaacagggatcga 3240  
OY 3241 tgcactgtgttattgtctgggggtcctcagaagggaaacctgtgaagccttatacagctc 3300  
Db 3241 tgcactgtgttattgtctgggggtcctcagaagggaaacctgtgaagccttatacagctc 3300  
OY 3301 atttatgtgctgtgaagaatctctcggagatgtgtgtacatattgaagcaagtgacttca 3360  
Db 3301 atttatgtgctgtgaagaatctctcggagatgtgtgtgtacatattgaagcaagtgacttca 3360  
OY 3361 gttgagggcagaatctctcggaaagaagctgtgaagcatctggaactccatgctgtgtag 3420  
Db 3361 gttgagggcagaatctctcggaaagaagctgtgaagcatctggaactccatgctgtgtag 3420  
OY 3421 tgtgtgggggttgggggtccctgggacactggctgtgtgaagggatctggcaaggacaca 3480  
Db 3421 tgtgtgggggttgggggtccctgggacactggctgtgtgaagggatctggcaaggacaca 3480  
OY 3481 gggcccttactgaaccatcagcatgtcagtlgcatltaaaagccatgcaagcttggagggc 3540  
Db 3481 gggcccttactgaaccatcagcatgtcagtlgcatltaaaagccatgcaagcttggagggc 3540  
OY 3541 cactggaatgtctctgtagatattactgaagaagcaacagaagaagggcatgtgagccc 3600  
Db 3541 cactggaatgtctctgtagatattactgaagaagcaacagaagaagggcatgtgagccc 3600  
OY 3601 ttgggctctctgggaaatgggaaatcagccaagagactggaaggaattactctaaggtc 3660  
Db 3601 ttgggctctctgggaaatgggaaatcagccaagagactggaaggaattactctaaggtc 3660  
OY 3661 agagaanaacaaagagatgtgtgtctctggaagcttgagcttcttataccaactcatc 3720  
Db 3661 agagaanaacaaagagatgtgtgtctctggaagcttgagcttcttataccaactcatc 3720  
OY 3721 ccttctccaataagccaactgtgtagttggccctccgggttgaaggcaagagagaga 3780  
Db 3721 ccttctccaataagccaactgtgtagttggccctccgggttgaaggcaagagagaga 3780  
OY 3781 aaggaacaggttctgggaacaaagacttctccgtgcaataagcctgggaaggaataaagga 3840  
Db 3781 aaggaacaggttctgggaacaaagacttctccgtgcaataagcctgggaaggaataaagga 3840  
OY 3841 tagagtgttaaaataaaaaaaanaa 3871  
Db 3841 tagagtgttaaaataaaaaaaanaa 3871  
|||||





Db 901 acctcaaccagttatagaggtatccggtltgagaaatttgytgcattatcatc 960  
QY 961 acctgaagtttgcttatataagttaagaccttgaaagatgtagccctctgcaagta 1020  
Db 961 acctgaagtttgcttatataagttaagaccttgaaagatgtagccctctgcaagta 1020  
QY 1021 gctatctgctcttgatcagaatgacaaagtcatacgacgaacccgtgttataccagt 1080  
Db 1021 gctatctgctcttgatcagaatgacaaagtcatacgacgaacccgtgttataccagt 1080  
QY 1081 accacagggagaccttggggcctctcgatgtagcaactccagcccagacctgagcaagttc 1140  
Db 1081 accacagggagaccttggggcctctcgatgtagcaactccagcccagacctgagcaagttc 1140  
QY 1141 agttctttaatgtgaccacaactccagaagagctgtatgacttctgtaaggaaatataa 1200  
Db 1141 agttctttaatgtgaccacaactccagaagagctgtatgacttctgtaaggaaatataa 1200  
QY 1201 tggatgatatgagggagagaagtgtggaatatgtgatatgacctcttggaacttgaggaga 1260  
Db 1201 tggatgatatgagggagagaagtgtggaatatgtgatatgacctcttggaacttgaggaga 1260  
QY 1261 ccagctagccccaagacaacaaagagactcctctctgctgcttcaaggaaacacagatcct 1320  
Db 1261 ccagctagccccaagacaacaaagagactcctctctgctgcttcaaggaaacacagatcct 1320  
QY 1321 tgtcctttcccaacagcccaagctgttatacctcagaagccttcttatactccaaa 1380  
Db 1321 tgtcctttcccaacagcccaagctgttatacctcagaagccttcttatactccaaa 1380  
QY 1381 gtgaaagggagagcccgctctctcttaactgcatgcatcaggggtgagccctgcttcc 1440  
Db 1381 gtgaaagggagagcccgctctctctcttaactgcatgcatcaggggtgagccctgcttcc 1440  
QY 1441 atcttcacacctgycacactcatgttccacactatcttccacctttttttgaga tgg 1500  
Db 1441 atcttcacacctgycacactcatgttccacactatcttccacctttttttgaga tgg 1500  
QY 1501 gctctgctcttctgcccagctgagtgacatgagacggttctcagctcaactgcaacctcc 1560  
Db 1501 gctctgctctcttctgcccagctgagtgacatgagacggttctcagctcaactgcaacctcc 1560  
QY 1561 gctctctgggttccaagaattctgctgcatcagccctccagagtaacttgagattacagca 1620  
Db 1561 gctctctgggttccaagaattctgctgcatcagccctccagagtaacttgagattacagca 1620  
QY 1621 tgtgccaacagcccggtcaatttctgattcttagtagagacgggttcttgcagatgtg 1680  
Db 1621 tgtgccaacagcccggtcaatttctgattcttagtagagacgggttcttgcagatgtg 1680  
QY 1681 ccaggtgtgtctcgaactcttgactcagatgatacactgtgcttggtcctccacaagtgc 1740  
Db 1681 ccaggtgtgtgtcgaactcttgactcagatgatacactgtgcttggtcctccacaagtgc 1740  
QY 1741 tgggattacaagcgltgagccacatgcccggcctctcttcacaccttacaactgtcttc 1800  
Db 1741 tgggattacaagcgltgagccacatgcccggcctctcttcacaccttacaactgtcttc 1800  
QY 1801 tatctcacaactgttttacaacactcactccctgtcttcctcaatgttccaaactgtcttc 1860  
Db 1801 tatctcacaactgttttacaacactcactccctgtcttcctcaatgttccaaactgtcttc 1860  
QY 1861 ccaatgttcaatgagctcttcttaccatttgggttgaaggagagttctctctgtctg 1920  
Db 1861 ccaatgttcaatgagctcttcttaccatttgggttgaaggagagttctctctgtctg 1920  
QY 1921 ttttttgttttcccaagaanaatcagttatatttttaataagaanaaatcattccagaa 1980  
Db 1921 ttttttgttttcccaagaanaatcagttatatttttaataagaanaaatcattccagaa 1980  
QY 1981 gatgataaattgtgaaanaacctcttctgctatttcttccagatttagtctccttc 2040  
Db 1981 gatgataaattgtgaaanaacctcttctgctatttcttccagatttagtctccttc 2040

QY 2041 tcccacccggagaaatgtgtggaagacataggctaaatttctccaggcctcacatgtctc 2100  
Db 2041 tcccacccggagaaatgtgtggaagacataggctaaatttctccaggcctcacatgtctc 2100  
QY 2101 ttcaactgtgtctgactgttaccactcttagcacccactgaaaacaaagtgtgtagagag 2160  
Db 2101 ttcaactgtgtctgactgttaccactcttagcacccactgaaaacaaagtgtgtagagag 2160  
QY 2161 tgtagagtgacagaatgttggtcttctgcccacttgcattccaaaattacaacggttgg 2220  
Db 2161 tgtagagtgacagaatgttggtcttctgcccacttgcattccaaaattacaacggttgg 2220  
QY 2221 ccgattccattttaggacaatgtctttagtataagttcccgagttctggaagagaagagc 2280  
Db 2221 ccgattccattttaggacaatgtctttagtataagttcccgagttctggaagagaagagc 2280  
QY 2281 cagagctgtctaatcttcatcattctcttcagtaataattatgtgactgactgtgtgc 2340  
Db 2281 cagagctgtctaatcttcatcattctcttcagtaataattatgtgactgactgtgtgc 2340  
QY 2341 taaggcatgacactgaggaaactagagatactcacagataacagggaaagtctccgtgtct 2400  
Db 2341 taaggcatgacactgaggaaactagagatactcacagataacagggaaagtctccgtgtct 2400  
QY 2401 catgaggtcttaacttcttacaaggagagaagatagccaataactatagaaatataatc 2460  
Db 2401 catgaggtcttaacttcttacaaggagagaagatagccaataactatagaaatataatc 2460  
QY 2461 aaggtatcatgtatgtatataatgtctgtgagaaanaataagccgggagagatgaaagaa 2520  
Db 2461 aaggtatcatgtatgtatataatgtctgtgagaaanaataagccgggagagatgaaagaa 2520  
QY 2521 tccctgagatgagagctgycagttttaaavtgggcccactgvggaaatgtgacgtttagcaga 2580  
Db 2521 tccctgagatgagagctgycagttttaaavtgggcccactgvggaaatgtgacgtttagcaga 2580  
QY 2581 gacgttagaggagatgagatctctgacaaaggatcccgagcagagaaagaaatgtgacag 2640  
Db 2581 gacgttagaggagatgagatctctgacaaaggatcccgagcagagaaagaaatgtgacag 2640  
QY 2641 ccccaagagtagaactgtctctagctggttcaaggaaagacagagagacccaaagagctgc 2700  
Db 2641 ccccaagagtagaactgtctctagctggttcaaggaaagacagagagacccaaagagctgc 2700  
QY 2701 tgggcaagggttagaatgtgaaagagagagcggtctgvggagagacaggtgtgtgagggccttg 2760  
Db 2701 tgggcaagggttagaatgtgaaagagagagcggtctgvggagagacaggtgtgtgagggccttg 2760  
QY 2761 ctctctgctaagttagagatgggaaacacttgtagaggttttgaaacagagaggtgcttgatgat 2820  
Db 2761 ctctctgctaagttagagatgggaaacacttgtagaggttttgaaacagagaggtgcttgatgat 2820  
QY 2821 ttatatcttgcagaaggttcaattctagctgtgcaataatgtgaaaaactttagtgcagaagggc 2880  
Db 2821 ttatatcttgcagaaggttcaattctagctgtgcaataatgtgaaaaactttagtgcagaagggc 2880  
QY 2881 agaaagaaagagagaaagacactgttagagagactcgcgaaggttccaggtcttggcctgggc 2940  
Db 2881 agaaagaaagagagaaagacactgttagagagactcgcgaaggttccaggtcttggcctgggc 2940  
QY 2941 cacaagcaacagacgtgtgtaaaatcatgattatctttaaagagacaaatagatattgc 3000  
Db 2941 cacaagcaacagacgtgtgtaaaatcatgattatctttaaagagacaaatagatattgc 3000  
QY 3001 tggagatttgaatgtgagatgttgaagaaagagatlaatgatagtacatlaagtttttg 3060  
Db 3001 tggagatttgaatgtgagatgttgaagaaagagatlaatgatagtacatlaagtttttg 3060  
QY 3061 gcttgaaatagcaggaagaaatgtgagatcactcgaataatagagaaatagtgctgggt 3120  
Db 3061 gcttgaaatagcaggaagaaatgtgagatcactcgaataatagagaaatagtgctgggt 3120





Db 2041 tccccatccggaaagatggtggaagacatagagctaaattctccagagctacacatggtc 2100  
 QY 2101 ttccattggtctgactctgtacaaattctagaccacacgtgaaanaaagaattgagtagaag 2160  
 Db 2101 ttccattggtctgactctgtacaaattctagaccacacgtgaaanaaagaattgagtagaag 2160  
 QY 2161 tgtagagtgacgaagaatgtggtctttgtgcccaactgtgcatccccaanaataacacgtgtg 2220  
 Db 2161 tgtagagtgacgaagaatgtggtctttgtgcccaactgtgcatccccaanaataacacgtgtg 2220  
 QY 2221 ccgattcccatcttgaggaacaatgcttagtataagttcccgagttggaagaagaagaagc 2280  
 Db 2221 ccgattcccatcttgaggaacaatgcttagtataagttcccgagttggaagaagaagaagc 2280  
 QY 2281 cagagctgtctagtttcaatcttccatccagtaaatattttagtaaccactgtgtgc 2340  
 Db 2281 cagagctgtctagtttcaatcttccatccagtaaatattttagtaaccactgtgtgc 2340  
 QY 2341 tagcatgaccttggaactagagatactcacagaataacaggaagttccctgtgct 2400  
 Db 2341 tagcatgaccttggaactagagatactcacagaataacaggaagttccctgtgct 2400  
 QY 2401 catggaacttaacattcacagggagaagatagccaataacataggaataataataac 2460  
 Db 2401 catggaacttaacattcacagggagaagatagccaataacataggaataataataac 2460  
 QY 2461 aaggtatcatctagtagtaaatctgtgtggaataaataaagaacagggaggaagtagaagaa 2520  
 Db 2461 aaggtatcatctagtagtaaatctgtgtggaataaataaagaacagggaggaagtagaagaa 2520  
 QY 2521 tcccgagatgagagctgcagacttcaaatgaggcctcacctgaggaaatgtacactgagcaga 2580  
 Db 2521 tcccgagatgagagctgcagacttcaaatgaggcctcacctgaggaaatgtacactgagcaga 2580  
 QY 2581 gaagcttgagggaagtgtgactctggaacaaggcattccagggcagaagaaatgtgacactg 2640  
 Db 2581 gaagcttgagggaagtgtgactctggaacaaggcattccagggcagaagaaatgtgacactg 2640  
 QY 2641 ccccaagaagtgaagaactgtgactctgagtgaggaagaagagcaggagacagaacagagtcg 2700  
 Db 2641 ccccaagaagtgaagaactgtgactctgagtgaggaagaagagcaggagacagaacagagtcg 2700  
 QY 2701 tgggcagaaggttagaatactggaagaagagcggctgaggaggaacaggtgtgaggggccttgg 2760  
 Db 2701 tgggcagaaggttagaatactggaagaagagcggctgaggaggaacaggtgtgaggggccttgg 2760  
 QY 2761 cttctgtcaagtgagatgagggaacacacgtggaggtttggaacaggaagagtgcttattgat 2820  
 Db 2761 cttctgtcaagtgagatgagggaacacacgtggaggtttggaacaggaagagtgcttattgat 2820  
 QY 2821 ttatatcttgcaagggatcatctgactgagcgaatatgtgaaaaaacttttagtgacaaggggc 2880  
 Db 2821 ttatatcttgcaagggatcatctgactgagcgaatatgtgaaaaaacttttagtgacaaggggc 2880  
 QY 2881 agaaaggaagaggaagacactgttaggaagctactgcaaggttccaggtcttgagcctgggc 2940  
 Db 2881 agaaaggaagaggaagacactgttaggaagctactgcaaggttccaggtcttgagcctgggc 2940  
 QY 2941 caccggcaacagcgatgtgtcaaatatctagattttttgaaaaagagccaatagatttgc 3000  
 Db 2941 caccggcaacagcgatgtgtcaaatatctagattttttgaaaaagagccaatagatttgc 3000  
 QY 3001 tgaagattggaatgtgagtgtaagaagaaggaagatgaatgaatgacatgaattttttg 3060  
 Db 3001 tgaagattggaatgtgagtgtaagaagaaggaagatgaatgaatgacatgaattttttg 3060  
 QY 3061 gctctgaatagcagaagaatgtgagttacacgttactgaaataggaaggaatgtggctgggt 3120  
 Db 3061 gctctgaatagcagaagaatgtgagttacacgttactgaaataggaaggaatgtggctgggt 3120  
 QY 3121 aagtaagaagattgtgtgcaaaagcagctgtctgtgtgttggaatgaggaggtctgtgcgca 3180  
 Db 3121 aagtaagaagattgtgtgcaaaagcagctgtctgtgtgttggaatgaggaggtctgtgcgca 3180

Db 3121 aagtaagaagattgtgtgcaaaagcagctgtctgtgtgttggaatgaggaggtctgtgcgca 3180  
 QY 3181 aatcaaaatgagaaattctctcaaggtcgaagtcgaagcagaagctcgagaagaagatctgaa 3240  
 Db 3181 aatcaaaatgagaaattctctcaaggtcgaagtcgaagcagaagctcgagaagaagatctgaa 3240  
 QY 3241 tgcacttggtttatgtgtgagggtgtctctcagaaggaacactgtgaaaagcctttacagt 3300  
 Db 3241 tgcacttggtttatgtgtgagggtgtctctcagaaggaacactgtgaaaagcctttacagt 3300  
 QY 3301 attatttggtctgtgagaaggtctctctggaaggtgtgtgtatcatttgaaggaaggtatctta 3360  
 Db 3301 attatttggtctgtgagaaggtctctctggaaggtgtgtgtatcatttgaaggaaggtatctta 3360  
 QY 3361 gttagaggaagctctctcgaagaagaagctgttaggcatactggaacactacatgctgtgtag 3420  
 Db 3361 gttagaggaagctctctcgaagaagaagctgttaggcatactggaacactacatgctgtgtag 3420  
 QY 3421 tgtgttgagggttgagggtgtcctggaacactgtgctgtggaaggatctggaaggacacaca 3480  
 Db 3421 tgtgttgagggttgagggtgtcctggaacactgtgctgtggaaggatctggaaggacacaca 3480  
 QY 3481 ggcgcacctactgaacacatcagcatgtcagtgtgcatttaagccaatgcaagctggaagggc 3540  
 Db 3481 ggcgcacctactgaacacatcagcatgtcagtgtgcatttaagccaatgcaagctggaagggc 3540  
 QY 3541 cactgagatgtctctgagatctactgagaagcacaagaagaagccaatgagatgagccc 3600  
 Db 3541 cactgagatgtctctgagatctactgagaagcacaagaagaagccaatgagatgagccc 3600  
 QY 3601 ttgggtctctctggaagaatgaggaaatcagccaagaagactggaagaaggttaacctaaagtc 3660  
 Db 3601 ttgggtctctctggaagaatgaggaaatcagccaagaagactggaagaaggttaacctaaagtc 3660  
 QY 3661 agaaagaaaccaaagagagtggtgtgtctcgaagctgtgagcttcttttcaacctctatc 3720  
 Db 3661 agaaagaaaccaaagagagtggtgtgtctcgaagctgtgagcttcttttcaacctctatc 3720  
 QY 3721 ccttctcacaataagccaactgtgtagttggccctccaggtgttgaaggaagaagagaga 3780  
 Db 3721 ccttctcacaataagccaactgtgtagttggccctccaggtgttgaaggaagaagagaga 3780  
 QY 3781 aaggaacagcgtttgggaagaagaactttctcgaataagcctggaagaagaataaagga 3840  
 Db 3781 aaggaacagcgtttgggaagaagaactttctcgaataagcctggaagaagaataaagga 3840  
 QY 3841 tagagtggttaaatataaaaaaataaaaaa 3871  
 Db 3841 tagagtggttaaatataaaaaaataaaaaa 3871

RESULT 4  
 AAF16225  
 ID AAF16225 standard; cDNA; 2549 BP.  
 XX  
 AC AAF16225;  
 XX  
 DT 13-MAR-2001 (first entry)  
 XX  
 DE Human prostate cancer antigen nucleotide sequence SFD ID NO:660.  
 XX  
 DE Human; prostate cancer; prostate cancer antigen; detection; diagnosis;  
 KW neoplastic; cytosolic; cardiovascular; immunomodulatory; muscular;  
 KW vulnerability; gastrointestinal; nephrotoxic; antiinfective; gynaecological;  
 KW antibacterial; gene therapy; neural; immune; reproductive; renal;  
 KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;  
 KW wound; infectious disease; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO20005174-A1.  
 XX  
 PD 21-SEP-2000.



[illegible]

RESULT	5	
AAH14525		
ID	AAH14525	standard; cDNA; 1661 BP.
XX		
AC	AAH14525;	
XX		
DT	26-JUN-2001	(first entry)
XX		
DE	Human cDNA sequence SEQ ID NO:12066.	
XX		
KW	Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss	
XX		
OS	Homo sapiens.	
XX		
PN	EP1074617-A2.	
XX		
PD	07-FEB-2001.	
XX		
PE	28-JUL-2000; 2000EP-0116126.	
XX		
PR	29-JUL-1999; 99JP-0248036.	
PR	27-AUG-1999; 99JP-0300253.	
PR	11-JAN-2000; 2000JP-0118776.	
PR	02-MAY-2000; 2000JP-0183767.	
PR	09-JUN-2000; 2000JP-0241899.	
XX		
PA	(HELI-) HELIX RES INST.	
XX		
PI	Oca T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;	
PI	Isihii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;	
XX		
DR	WPI; 2001-318749/34.	
XX		
PT	Primer sets for synthesizing polynucleotides, particularly the 5602	
PT	full-length cDNAs defined in the specification, and for the detection	
PT	and/or diagnosis of the abnormality of the proteins encoded by the	
PT	full-length cDNAs -	

xx Claim 8; SEQ ID 12066; 2537bp + CD ROM; English.

xx

xx The present invention describes primer sets for synthesising 5602

CC full-length cDNAs defined in the specification. Where a primer set

CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary

CC to the complementary strand of a polynucleotide which comprises one of

CC the 5602 nucleotide sequences defined in the specification, where the

CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination

CC of an oligonucleotide comprising a sequence complementary to the

CC complementary strand of a polynucleotide which comprises a 5'-end

CC sequence and an oligonucleotide comprising a sequence complementary to a

CC polynucleotide which comprises a 3'-end sequence, where the

CC oligonucleotide comprises at least 15 nucleotides and the combination of

CC the 5'-end sequence/3'-end sequence is selected from those defined in

CC the specification. The primer sets can be used in antisense therapy and

CC in gene therapy. The primers are useful for synthesising polynucleotides

CC particularly full-length cDNAs. The primers are also useful for the

CC detection and/or diagnosis of the abnormality of the proteins encoded by

CC the full-length cDNAs. The primers allow obtaining of the full-length

CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and

CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to

CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632

CC represent oligonucleotides, all of which are used in the exemplification

CC of the present invention.

xx

xx Sequence 1961 BP; 417 A; 579 C; 481 G; 484 T; 0 other;

[illegible]



OY	624	taaagagccgctgctgtgcccagggactaatataaagacctggaaccagatcatatga	683
Db	601	tataagagccgctcgctggtccagaggaataataaagacctggaaccagatcatatga	660
OY	684	agccctgtcatccagagacgtgcggtgacatacaacggtgagaaactggaacatccatcga	743
Db	661	agccctgtcatccagagacgtgcggtgacatacaacggtgagaaactggaacatccatcga	720
OY	744	gacatgagctgtgacctcttccgaactcttccaaaagcctttaaagatgtctcgaacctgc	803
Db	721	gacatgagctgtgacctcttccgaactcttccaaaagcctttaaagatgtctcgaacctgc	780
OY	804	gagaggtccagggagatccaaaggactccaaagattcttactcttccatagcaatatat	863
Db	781	gagaggtccagggagatccaaaggactccaaagattcttactcttccatagcaatatat	840
OY	864	gtaaagatctctggaatgcacaaataaagctgtgaaagagaacctaccccaattatagagcg	923
Db	841	gtaaagatctctggaatgcacaaataaagctgtgaaagagaacctaccccaattatagagcg	900
OY	924	tatccggtctgagaataattgtgctaacatgtaacatctactgtcaagtttgctattataag	983
Db	901	tatccggtctgagaataattgtgctaacatgtaacatctactgtcaagtttgctattataag	960
OY	984	ttgaaagaccttgaaagaattgtgaagccctctgtgcacgtatcatctgctctttgataaag	1043
Db	961	ttgaaagaccttgaaagaattgtgaagccctctgtgcacgtatcatctgctctttgataaag	1020
OY	1044	gacaaagtcatcagcagacgaacccctgtgcatatacaatccaaaggagacacttggggccct	1103
Db	1021	gacaaagtcatcagcagacgaacccctgtgcatatacaatccaaaggagacacttggggccct	1080
OY	1104	tcgaaatgagaccttccagcccagaccttgaaagcagtttcaattttaaattgacacactc	1163
Db	1081	tcgaaatgagaccttccagcccagaccttgaaagcagtttcaattttaaattgacacactc	1140
OY	1164	cagaagagagctgtatgaacttctgtaaggaaataataatgatatgatatgagaggagaagt	1223
Db	1141	cagaagagagctgtatgaacttctgtaaggaaataataatgatatgatatgagaggagaagt	1200
OY	1224	gtggaataatcttggaatgagccctctggaaactggaggagaccagttagcccaagcaacaa	1283
Db	1201	gtggaataatcttggaatgagccctctggaaactggaggagaccagttagcccaagcaacaa	1260
OY	1284	gagagctccctctggtgcgtcttcaagaaacacagatcttcttcttccaaagacggccgg	1343
Db	1261	gagagctccctctggtgcgtcttcaagaaacacagatcttcttcttccaaagacggccgg	1320
OY	1344	ctgttgaataccccaagacctctctcttactctccaagtgaagggaagccccgctctct	1403
Db	1321	ctgttgaataccccaagacctctctcttactctccaagtgaagggaagccccgctctct	1380
OY	1404	ctaaactgcagtatacaagggtgtagcgtctcttctactccaactgcacactgcactatg	1463
Db	1381	ctaaactgcagtatacaagggtgtagcgtctcttctactccaactgcacactgcactatg	1440
OY	1464	ttcacacactactcttccacacttttttttggagatgagactgcgtctcttgcacagagtg	1523
Db	1441	ttcacacactactcttccacacttttttttggagatgagactgcgtctcttgcacagagtg	1500
OY	1524	gagtgcaatgagcagcttctccaagctcaactgcacacctccgcctcttgggtttcaagcaatct	1583
Db	1501	gagtgcaatgagcagcttctccaagctcaactgcacacctccgcctcttgggtttcaagcaatct	1560
OY	1584	gctgtcatcaagccctcccgagatcactgtggaattaaagcagtgcacacacgcccgttaatt	1643
Db	1561	gctgtcatcaagccctcccgagatcactgtggaattaaagcagtgcacacacgcccgttaatt	1620
OY	1644	tttgtatcttttaagacagagacgggttttctgcacagtgtgccaagcgctgtctcgaaccttga	1703
Db	1621	tttgtatcttttaagacagagacgggttttctgcacagtgtgccaagcgctgtctcgaaccttga	1680
OY	1704	cttcaagaatgatacatctgtccttgcctccccaagatgcttggaattacaagcgctlgagccacc	1763

Db	1681	cttcaagatgattcattctgcttgcttgcctccaaagtcgggagttcaagggtggccacc	1740
QY	1764	atgcccgagctctttctcaaccttaacacctgtctcttatccccaacatcgtttcacac	1823
Db	1741	atgcccgagctctttctcaaccttaacacctgtctcttatccccaacatcgtttcacac	1800
QY	1824	cttcatcccgctctctccatgttcaacacttgctctccccaagtccatagctgccttct	1883
Db	1801	cttcatcccgctctctccatgttcaacacttgctctccccaagtccatagctgccttct	1860
QY	1884	taccatttgggtttaaggcagctctctcttgctctgtttttgtttttccagaatat	1943
Db	1861	taccatttgggtttaaggcagctctctctcttgctctgtttttgtttttccagaatat	1920
QY	1944	cagtatatttttttaataagaanaaacattccctagaagatg	1984
Db	1921	cagtatatttttttaataagaanaaacattccctagaagatg	1961

## RESULT 6

AAH31138  
ID AAH31138 standard; cDNA; 1586 BP.  
XX  
AC AAH31138;  
XX  
XX 27-JUL-2001 (first entry)  
DT  
XX  
DE Human colon cancer cell line Km12L4-A cDNA library derived seq#1072.  
XX  
KM Human; diagnosis; colon cancer; cancer; malignant; chromosome mapping;  
KM detection; colon cancer cell line Km12L4-A; ss.  
XX  
OS Homo sapiens.  
OS  
XX  
PN WO200018916-A2.  
XX  
PD 06-APR-2000.  
XX  
PF 23-SEP-1999; 99WO-US222226.  
XX  
XX 28-SEP-1998; 98US-0102161.  
PR 28-SEP-1998; 98US-0102180.  
PR 29-SEP-1998; 98US-0102380.  
PR 08-OCT-1998; 98US-0103815.  
PR 27-OCT-1998; 98US-0105877.  
XX  
PA (CHIR ) CHIRON CORP.  
PA (HYSE ) HYSEQ INC.  
PI Williams LT, Escobedo J, Innis MA, Garcia PD, Suduth-Klinger J;  
PI Reinhard C, Gliese K, Randazzo F, Kennedy GC, Pot D, Kassam A;  
PI Lanson G, Drmanac S, Cirvenjakov R, Dickson M, Drmanac S, Labat I;  
PI Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;  
XX  
DR WPI: 2000-293155/25.  
XX  
XX Polynucleotide library comprising 1079 defined sequences, useful in  
PT the form of an array to detect cancer or susceptibility to cancer -  
PS  
PS Claim 1; Page 498-499; 502pp; English.  
XX  
XX The present invention describes a library of polynucleotides comprising  
CC 1079 nucleotide sequences (given in AAH30067 to AAH31145). Also described  
CC are: (1) an isolated polynucleotide (I) having at least 90% identity to  
CC one of the 1079 sequences; (2) a recombinant host cell containing (1);  
CC (3) an isolated polypeptide (II) encoded by (1); (4) an antibody that  
CC specifically binds to (II); (5) a vector comprising (I); and (6) a method  
CC of detecting differentially expressed genes correlated with a cancerous  
CC state of a mammalian cell comprising detecting a gene product encoded by  
CC 65 of the 1079 sequences given in the specification. The polynucleotides  
CC are used to monitor patients having (or susceptible) to cancer to detect  
CC potentially malignant events at a molecular level before they are

CC detectable at a gross morphological level. The polynucleotides are also useful for monitoring the efficacy of various therapies and preventive interventions. Polynucleotide probes based on the disclosed sequences CC are useful for chromosome mapping and detection of transcription levels. CC The 1079 polynucleotide sequences were derived from a human colon cancer cell line Km12L4-A CDNA library.

XX Sequence 1586 BP; 381 A; 406 C; 340 G; 453 T; 6 other;

Query Match 38.1%; Score 1474.4; DB 21; Length 1586;  
Best Local Similarity 98.3%; Pred. No. 1,3e-287;  
Matches 1506; Conservative 5; Mismatches 16; Indels 5; Gaps 2;

QY 559 ccgcgtcacacaccttactgaagcattccctgatacgaacatgataagaggaacatg 618  
Db 1 ccgctgcccacaccttactgaagcattccctgatacgaacatgataagaggaacatg 60  
QY 619 catattaaagagcctctgctgctgcgaagactatacattaaagacctggaacccaagtcac 678  
Db 61 catattaaagagcctctgctgctgcgaagactatacattaaagacctggaacccaagtcac 120  
QY 679 atgaagagcttataccgaagcagtgccggaatacaacggtgagaacatgaggaacatcca 738  
Db 121 atgaagagcctgtcaatccgaagcagtgccggaatacaacggtgagaacatgaggaacatcca 180  
QY 739 tcacagacatgagcctgagccctcccgactcttcaaacgctttacagagtcctgcag 798  
Db 181 tcacagacatgagcctgagccctcccgactcttcaaacgctttacagagtcctgcag 240  
QY 799 cctcgaagagcttcaagagatcaagagatcaagagatttcaaccttccataagacatc 858  
Db 241 cctcgaagagcttcaagagatcaagagatcaagagatttcaaccttccataagacatc 300  
QY 859 attatgtgaagcttctgtaagcaaatatacagtgtaagaaagaaaccccaagttatag 918  
Db 301 attatgtgaagcttctgtaagcaaatatacagtgtaagaaagaaaccccaagttatag 360  
QY 919 gaggcataccggttgaagaatttctgctacatgatacattacttgaagttgctcatt 978  
Db 361 gaggcataccggttgaagaatttctgctacatgatacattacttgaagttgctcatt 420  
QY 979 ataagttgaagcacttgaagaatcagccctctgcaagtcagtcattctgcttctgac 1038  
Db 421 ataagttgaagcacttgaagaatcagccctctgcaagtcagtcattctgcttctgac 480  
QY 1039 agaattgaagagctatgacagagacactggtatattaccagttaccacagggagacacttgg 1098  
Db 481 agaattgaagagctatgacagagacactggtatattaccagttaccacagggagacacttgg 540  
QY 1099 gacctcggatgacacttccagcccaagcactgaagcagttcagttcctttaatgtgacca 1158  
Db 541 gacctcggatgacacttccagcccaagcactgaagcagttcagttcctttaatgtgacca 600  
QY 1159 cactccgaagaagagctgatacttctgtaagaaataataatgataatgagggag 1218  
Db 601 cactccgaagaagagctgatacttctgtaagaaataataatgataatgagggag 660  
QY 1219 aagttgtgaatatgtgaatgacactctgtgaacatgagaagaccgacgacccaagaa 1278  
Db 661 aagttgtgaatatgtgaatgacactctgtgaacatgagaagaccgacgacccaagaa 720  
QY 1279 ccaaaagagacttctctgtgcttcaagaaacacagattccttctccttcccaacagc 1338  
Db 721 ccaaaagagacttctctgtgcttcaagaaacacagattccttctccttcccaacagc 780  
QY 1339 ccaggcgtgtgatacctcaagagccttcttactctccaaagtgaagggagccccc 1398  
Db 781 ccaggcgtgtgatacctcaagagccttcttactctccaaagtgaagggagccccc 840  
QY 1399 tctcttaacagcattgatacaggggtgagcctgcttcttacttaccacctggccac 1458  
Db 841 tctcttaacagcattgatacaggggtgagcctgcttcttacttaccacctggccac 900

QY 1459 tcaattcacacatacttctcacaacttttttgaagatgagatcgcctctcccca 1518  
Db 901 tcaattcacacatacttctcacaacttttttgaagatgagatcgcctctcccca 960  
QY 1519 ggcctggaatgcaatgacagcttctcagctacatgcaacctccgcctcttggatcaaga 1578  
Db 961 ggcctggaatgcaatgacagcttctcagctacatgcaacctccgcctcttggatcaaga 1020  
QY 1579 attcctgatacagccctccagatcctggaatgatacagagatgacacagcccgagc 1638  
Db 1021 attcctgatacagccctccagatcctggaatgatacagagatgacacagcccgagc 1080  
QY 1639 taatttgaatttttaagtagagagagcgggttttgcattgttggccagcgtgtctcgaact 1698  
Db 1081 taatttgaatttttaagtagagagagcgggttttgcattgttggccagcgtgtctcgaact 1140  
QY 1659 ctgacttcagatgatacctatgcttgccttgcctcccaagctgctggatataagcgtgag 1738  
Db 1141 ctgacttcagatgatacctatgcttgccttgcctcccaagctgctggatataagcgtgag 1200  
QY 1759 ccacatgcccggcctcttctcacttaccactgacttcttacttccacatcgttct 1818  
Db 1201 ccacatgcccggcctcttctcacttaccactgacttcttacttccacatcgttct 1260  
QY 1819 cacacattcacctcctgtcttctccatgttccacacttgccttcccaatgttcatagctgac 1878  
Db 1261 cacacattcacctcctgtcttctccatgttccacacttgccttcccaatgttcatagctgac 1320  
QY 1879 ttcttaccatttgggttgaagggcagcttctcctgctgcttcttcttcttctccag 1938  
Db 1321 ttcttaccatttgggttgaagggcagcttctcctgctgcttcttcttcttcttctccag 1380  
QY 1939 aaatcacatattatttttaataaagaanaaacattccacagaagatgataatgtgaaac 1998  
Db 1381 aaatcacatattatttttaataaagaanaaacattccacagaagatgataatgtgaaac 1440  
QY 1999 ctcccttggcttatttgccttccag--atttagctccttcttcccaatccgggaaag 2056  
Db 1441 ctcccttggcttatttgccttccagattttaagctccttcttcccaatccgggaa 1500  
QY 2057 a---tggtggaagacatagactaaattctcc 2085  
Db 1501 agatgggtggaagacatagactaaattctcc 1532

# RESULT 7

AAH06808  
ID AAH06808 standard; CDNA: 818 BP.

XX AAH06808;

DT 26-JUN-2001 (first entry)

XX

DE Human cDNA clone (5'-primer) SEQ ID NO:3643.

XX

KW Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.

OS Homo sapiens.

XX

FN EPI074617-A2.

XX

PD 07-FEB-2001.

XX

PF 28-JUL-2000; 2000EP-0116126.

XX

PR 29-JUL-1999; 99JP-0248036.

PR 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.

XX

PA (HELI-) HELIX RES INST.



SO Sequence 618 BP; 157 A; 133 C; 169 G; 136 T; 23 other:

Query Match 12.0%; Score 466.4; DB 21; Length 618;  
Best Local Similarity 94.8%; Pred. No. 1.2e-84;  
Matches 506; Conservative 0; Mismatches 25; Indels 3; Gaps 3;

```
Oy 1076 ccaatcacacagagacacttgaggcctcgcgagatgagacattccagccacagactgaagc 1135
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 534 CNAGTCCNNAGGCCANTGGGCTTNGATGAGCANTTCCAGCCCACTTAAAC 475

Oy 1136 agttcagttcttaattgagcca-cactccagaagaagc-gtattgattt-gctaaaga 1192
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 474 AGTTGAGTTCTTAAATGAGNCCACCCTCCAGANAGAGCTGTATGACTTTGGNTAAGGA 415

Oy 1193 aaataaagatgatgatgatgagaggaagtgtggaataatgtagatgagcccttgaaact 1252
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 414 AAATATATGATGATGATGAGGAGAGTGTGSAATATGATGATGACCTTTGAACT 355

Oy 1253 ggaagagacagcctagccacagaacaaagaagactcctcttgcgctcagaagaac 1312
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 354 GGAGAGAGACAGTTAGCCACAGCAACCAAGAGACTTCTTGGCGCTCAGGAAACAC 295

Oy 1313 agattcttgccctttcccaagagccagcgtgtgattactaagagccttcttcttac 1372
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 294 AGATTCCTTGGCTTTCCTCCAAACAGCCAGCTGTGTATACCTCAGAGCTTCTTTTAC 235

Oy 1373 tctccaaagtgaagagagagcccgctctccttaactcatgcatcaggggtgagcctg 1432
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 234 TCTCCAAAGTGAAGGAGAGCCCGCTGTCTCTAATCTCATGTCATGAGGGGTAGCCTG 175

Oy 1433 ccttctctatctcaacactgccaactcaatgltcaacactatcttctacatttttt 1492
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 174 CTTTCCATCTTACACACTCCACCTCATCTGTACACCTTCTTCTTCTTCTTCTTCTTCTT 115

Oy 1493 gagatgagctcgtcctccttgccagagctgagtgcaatgcaatgcaatgcaatgca 1552
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 114 GAGATGAGGCTCGCTCTCTTGGCCAGGCTGAGAGTGCACGCTTCTCAGCTCAGCTG 55

Oy 1553 caactccgcctcttggtgtcaagaacttctgctgcatcagccctccagatgac 1606
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 54 CAACCTCGGCTCTTGGGTTCAGCAATTCGTGATCAGACCTCCGCAAGTAC 1
```

```
RESULT 9
AAH30585
ID AAH30585 standard; cDNA; 408 BP.
XX
AC AAH30585;
XX
DT 27-JUL-2001 (first entry)
XX
DE Human colon cancer cell line Km12L4-A cDNA library derived sequence #519.
XX
KW Human; diagnosis; colon cancer; cancer; malignant; chromosome mapping;
XX detection; colon cancer cell line Km12L4-A; ss.
XX
OS Homo sapiens.
XX
PN MO200018916-A2.
XX
PD 06-APR-2000.
XX
PE 23-SEP-1999; 99WO-US22226.
XX
PR 28-SEP-1998; 98US-0102161.
XX 28-SEP-1998; 98US-0102180.
XX 29-SEP-1998; 98US-0102380.
XX 08-OCT-1998; 98US-0103815.
XX 27-OCT-1998; 98US-0105877.
XX
PA (CHIR ) CHIRON CORP.
XX (HISE-) HYSED INC.
```

```
XX Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
PI Lamson G, Drmanac R, Crivenjakov R, Dickson M, Drmanac S, Labat I;
PI Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;
XX WPI: 2000-293155/25.
XX
PT Polynucleotide library comprising 1079 defined sequences, useful in
PS the form of an array to detect cancer or susceptibility to cancer -
XX Claim 1; Page 333; 502pp; English.
```

The present invention describes a library of polynucleotides comprising 1079 nucleotide sequences (given in AAH30067 to AAH31145). Also described are: (1) an isolated polynucleotide (I) having at least 90% identity to one of the 1079 sequences; (2) a recombinant host cell containing (1); (3) an isolated polypeptide (II) encoded by (1); (4) an antibody that specifically binds to (II); (5) a vector comprising (1); and (6) a method of detecting differentially expressed genes correlated with a cancerous state of a mammalian cell comprising detecting a gene product encoded by 65 of the 1079 sequences given in the specification. The polynucleotides are used to monitor patients having (or susceptible) to cancer to detect potentially malignant events at a molecular level before they are detectable at a gross morphological level. The polynucleotides are also useful for monitoring the efficacy of various therapies and preventive interventions. Polynucleotide probes based on the disclosed sequences are useful for chromosome mapping and detection of transcription levels. The 1079 polynucleotide sequences were derived from a human colon cancer cell line Km12L4-A cDNA library.

Sequence 408 BP; 117 A; 96 C; 98 G; 97 T; 0 other:

Query Match 10.5%; Score 406.4; DB 21; Length 408;  
Best Local Similarity 99.8%; Pred. No. 1.3e-72;  
Matches 407; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Oy 559 ccgctgtcacaccttctactgaagcctcctgaatgaagaaatgtaagaagaacatg 618
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 ccgctgtcacaccttctactgaagcctcctgaatgaagaaatgtaagaagaacatg 60

Oy 619 catatttaagaagcctgctggtgcgagagctacattaaagactggaacaaatgcat 678
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 catatttaagaagcctgctggtgcgagagctacattaaagactggaacaaatgcat 120

Oy 679 atgaagcctgttcatccagacagtgcgagatacaacggtgaagaactggaacatcca 738
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 atgaagcctgttcatccagacagtgcgagatacaacggtgaagaactggaacatcca 180

Oy 739 tcacagacatgagctgagccctccgactcttcaagaaccttctacagatgctcgag 798
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 181 tcacagacatgagctgagccctccgactcttcaagaaccttctacagatgctcgag 240

Oy 799 cctgcgaggttccagaggaatcaagaactcaagaattctacattccatagcagatc 858
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 241 cctgcgaggttccagaggaatcaagaactcaagaattctacattccatagcagatc 300

Oy 859 attatgtagaagctctggaatgaagaataacagtggtgaagaagaacctcaaccagttatag 918
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 301 attatgtagaagctctggaatgaagaataacagtggtgaagaagaacctcaaccagttatag 360

Oy 919 gagctatccggttgagaataattgtgctacacatgatactattctgac 966
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 361 gagctatccggttgagaataattgtgctacacatgatactattctgac 408
```

```
RESULT 10
AAH50786
ID AAH50786 standard; cDNA; 401 BP.
XX
AC AAH50786;
XX
```

DT 23-AUG-2001 (first entry)  
 XX Human tumour associated CDNA #115.  
 DE  
 XX Human: cancer specific gene expression; gene therapy;  
 KW age related differential expression; ss.  
 KW  
 OS Homo sapiens.  
 XX  
 XX W0200136685-A2.  
 XX  
 PD 25-MAY-2001.  
 XX  
 PF 17-NOV-2000; 2000MO-US31809.  
 XX  
 PR 17-NOV-1999; 99US-0166056.  
 PR 17-NOV-1999; 99US-0166106.  
 XX  
 PA (NYXI-) NYXIS NEURO THERAPIES INC.  
 XX  
 PI Kroes RA, Moskal JR, Yamamoto H;  
 XX  
 DR WPI; 2001-355647/37.  
 XX  
 PT Novel nucleic acid molecules differentially expressed in brain cancers,  
 PT useful for ascertaining propensity of cell for malignant phenotype or  
 PT ascertaining suitability of anti-neoplastic drug candidate -  
 XX  
 PS Claim 26; Page 53; 82pp; English.  
 XX  
 CC The present invention provides the sequences of 184 CDNA fragments which  
 CC are differentially expressed in cancer cell depending on the age of the  
 CC patient. They can be used to diagnose and identify treatments for  
 CC cancers, particularly brain cancers such as haemangioblastoma, teratoma,  
 CC haemangioma, glioblastoma, schwannoma, osteoma and pinealoma. The  
 CC present sequence is a cancer-associated CDNA of the invention.  
 XX  
 XX  
 SQ Sequence 401 BP; 133 A; 72 C; 90 G; 106 T; 0 other;

Query Match 9.6%; Score 371.6; DR 22; Length 401;  
 Best Local Similarity 97.7%; Pred. No. 1.3e-65;  
 Matches 377; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

OY 2129 agacccacgtaaaagttgtagagagtgtagagtcgagaaatggtgttggcc 2188  
 |||||||  
 DB 14 agacccacgtaaaagttgtagagagtcgtagagtcgagaaatggtgttggcc 73  
 OY 2189 ccacttgcacatccaaatatacaacggttgcgcgcatccattgaggaacatgctagt 2248  
 |||||||  
 DB 74 ccacttgcacatccaaatatacaacggttgcgcgcatccattgaggaacatgctagt 133  
 OY 2249 tataagttccgagttggaagaagaagccagagctgtctagtcttcaattctt 2308  
 |||||||  
 DB 134 tataagttccgagttggaagaagaagccagagctgtctagtcttcaattctt 193  
 OY 2309 cagtaatatattttagtactactctgtgtgtgtaggcatgagcttggaactagagatac 2368  
 |||||||  
 DB 194 cagtaatatattttagtactactctgtgtgtgtaggcatgagcttggaactagagatac 233  
 OY 2369 ttacacgaataacaggaaggttcctgtgtctcatgtgaggttcaattctacagggagaa 2428  
 |||||||  
 DB 254 ttacacgaataacaggaaggttcctgtgtctcatgtgaggttcaattctacagggagaa 313  
 OY 2429 gagatggccaatacatagaataataataatacaaggtatcatgtagtataattgtctgtg 2488  
 |||||||  
 DB 314 gagatggccaatacatagaataataataatacaaggtatcatgtagtataattgtctgtg 373  
 OY 2489 gagaaaaataagcaggaggagagt 2514  
 |||||||  
 DB 374 gaaaaaaaagcttaaggcggaat 399

RESULT 11  
 ID AAT21281  
 XX AAT21281 standard; CDNA to mRNA; 400 BP.  
 AC  
 XX AAT21281;  
 XX  
 DT 15-AUG-1996 (first entry)  
 XX  
 DE Human gene signature HUMGS02637.  
 XX  
 KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
 KW human; cloning; mapping; non-biased library; diagnosis; detection;  
 KW cell typing; abnormal cell function; ss.  
 XX  
 OS Homo. sapiens.  
 XX  
 PN W09514772-A1.  
 XX  
 PD 01-JUN-1995.  
 XX  
 PF 11-NOV-1994; 94MO-JP01916.  
 XX  
 PR 12-NOV-1993; 93JP-0355504.  
 XX  
 PA (MATS/) MATSUBARA K.  
 PA (OKUBO/) OKUBO K.  
 XX  
 PI Matsubara K, Okubo K;  
 XX  
 DR WPI; 1995-206931/27.  
 XX  
 PT Identifying gene signatures in 3'-directed human CDNA library - e.g.  
 PT for diagnosis of abnormal cell function, by preparing CDNA that  
 PT reflects relative abundance of corresp. mRNA in specific human  
 PT tissues  
 XX  
 PS Claim 1; Page 844; 2245pp; Japanese.  
 XX  
 CC A single-stranded DNA (or its complementary strand or the corresp.  
 CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
 CC given in AAT19001-T26837 and which is able to hybridise to part of  
 CC human genomic DNA, CDNA or mRNA is claimed. The GS (Gene Signature)  
 CC sequences were obtained from 3'-directed CDNA libraries prepared  
 CC from various human tissues; synthesis of CDNA was initiated from the  
 CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
 CC untranslated sequence is unique to a particular mRNA species, almost  
 CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
 CC is constructed so as to reflect accurately the relative abundance of  
 CC different mRNAs in the particular tissue from which it was derived.  
 CC The appearance frequency of a given GS in a CDNA library can be  
 CC determined (esp. using primers and probes derived from the GS  
 CC sequences) as a means of diagnosing abnormal cell function or for  
 CC recognising different cell types.  
 XX  
 XX  
 SQ Sequence 400 BP; 116 A; 82 C; 116 G; 85 T; 1 other;

Query Match 9.5%; Score 369.4; DR 16; Length 400;  
 Best Local Similarity 96.9%; Pred. No. 3.7e-65;  
 Matches 376; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

OY 3462 gatctggcaggagaccacagcccccactactgacacacagcagctgtaggttaaa 3521  
 |||||||  
 DB 1 gatctggcaggagaccacagcccccactactgacacacagcagctgtaggttaaa 60  
 OY 3522 gccatgcagctggaggggccactgagatgtctctgtgattactctgagaagcacaagaaa 3581  
 |||||||  
 DB 61 gccatgcagctggaggggccactgagatgtctctgtgattactctgagaagcacaagaaa 120  
 OY 3582 agagccatgatggagcccttggctctctgggaatgggaatcagccaagagctgag 3641  
 |||||||  
 DB 121 agagccatgatggagcccttggctctctgggaatgggaatcagccaagagctgag 180

QY 3642 aaggagttactcttaagtccagagaaacccaagagatgtgtgtctcttggaagcttagctt 3701  
DB 181 aaggagttactcttaagtccagagaaacccaagagatgtgtgtctcttggaagcttagctt 240  
QY 3702 tctttatccaacctcattccctctcccaataagccactgtgttagttgggcccctcag 3761  
DB 241 tctttatccaacctcattccctctcccaataagccactgtgttagttgggcccctcag 300  
QY 3762 ggttgaagcagaagagaaagcagcagcttgggaaacaagactttccctgcaatagc 3821  
DB 301 ggttgaagcagaagagaaagcagcagcagcttgggaaacaagactttccctgcaatagc 360  
QY 3822 ctgggaaggaataaagatagatgtgt 3849  
DB 361 ctgggaaggaataaangataagagt 388

RESULT 12  
AAH50808  
ID AAH50808 standard; CDNA: 407 BP.  
XX  
AC AAH50808;  
XX  
DT 23-AUG-2001 (first entry)  
XX  
DE Human tumour associated CDNA #137.  
XX  
KW Human; cancer specific gene expression; gene therapy;  
KM age related differential expression; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200136685-A2.  
XX  
PD 25-MAY-2001.  
XX  
PE 17-NOV-2000; 2000WO-US31809.  
XX  
PR 17-NOV-1999; 99US-0166056.  
PR 17-NOV-1999; 99US-0166106.  
XX  
PA (NXXI-) NYXIS NEURO THERAPIES INC.  
XX  
PI Kroes RA, Moskal JR, Yamamoto H;  
XX  
DR WPI: 2001-355647/37.  
XX  
PT Novel nucleic acid molecules differentially expressed in brain cancers,  
PT useful for ascertaining propensity of cell for malignant phenotype or  
PT ascertaining suitability of anti-neoplastic drug candidate -  
XX  
PS Claim 28; Page 57; 82pp; English.  
XX  
CC The present invention provides the sequences of 184 CDNA fragments which  
CC are differentially expressed in cancer cell depending on the age of the  
CC patient. They can be used to diagnose and identify treatments for  
CC cancers, particularly brain cancers such as haemangioblastoma, teratoma,  
CC haemangioma, glioblastoma, schwannoma, osteoma and pinealoma. The  
CC present sequence is a cancer-associated CDNA of the invention.  
XX  
XX  
SQ Sequence 407 BP; 136 A; 73 C; 91 G; 107 T; 0 other;

Query Match 9.2%; Score 355.6; DB 22; Length 407;  
Best Local Similarity 96.2%; Pred. No. 2.2e-62;  
Matches 377; Conservative 0; Mismatches 9; Indels 6; Gaps 1;

QY 2129 agcaccacgtaaaacaagtgtgtagagatgagatgcagaaatgtggtcttgc 2188  
DB 14 agcaccacgtaaaacaagtgtgtagagatgagatgcagaaatgtggtcttgc 73  
QY 2189 ccatttgcattcccaataatagaacggttgccgattcccttggagacaatgttagt 2248  
|||||

DB 74 ccatttgcattcccaataatagaacggttgccgattcccatcttgagacaatgttagt 133  
QY 2249 tataagttccgagttggaagaaagcagaagctgtctagttcatcttctt 2308  
DB 134 tataagttccgagttggaagaaagcagaagctgtctagttcatcttctt 193  
QY 2309 cagtaaatattattagttacactctgtgtctaggaattgacctgg-----gaactag 2362  
DB 194 cagtaaatattattagttacactctgtgtctaggaattgacctgggaactag 253  
QY 2363 agatacttcacagaataacaggaagttccctgtgctccatgagacttaccatccag 2422  
DB 254 agatacttcacagaataacaggaagttccctgtgctccatgagacttaccatccag 313  
QY 2423 gagaagaagatagccaatacataggaataataataacaaggtatcatgtatgtaatt 2482  
DB 314 gagaagaagatagccaatacataggaataataataacaaggtatcatgtatgtaatt 373  
QY 2483 gctgtggaagaaataaagcagggagggagt 2514  
DB 374 gctgtggaagaaataaagcagggagggagt 405

RESULT 13  
AAH1840/c  
ID AAH1840 standard; CDNA: 351 BP.  
XX  
AC AAH1840;  
XX  
DT 26-JUN-2001 (first entry)  
XX  
DE Human CDNA clone (3'-primer) SEQ ID NO:8675.  
XX  
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
XX  
OS Homo sapiens.  
XX  
PN EP1074617-A2.  
XX  
PD 07-FEB-2001.  
XX  
PE 28-JUL-2000; 2000EP-0116126.  
XX  
PR 29-JUL-1999; 99JP-0248036.  
PR 27-AUG-1999; 99JP-0300253.  
PR 11-JAN-2000; 2000JP-0118776.  
PR 02-MAY-2000; 2000JP-0183767.  
PR 09-JUN-2000; 2000JP-0241899.  
XX  
PA (HELI-) HELIX RES INST.  
XX  
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX  
DR WPI: 2001-318749/34.  
XX  
PT Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length CDNA defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
PT full-length CDNA -  
XX  
XX  
PS Claim 3; SEQ ID 8675; 2537pp + CD ROM; English.  
XX  
CC The present invention describes primer sets for synthesizing 5602  
CC full-length CDNA defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dt primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification; where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the



CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
CC AAB95893 represent human amino acid sequences; and AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.

SO Sequence 351 BP; 127 A; 60 C; 86 G; 67 T; 11 other;

Query Match 8.8%; Score 340; DB 22; Length 351;  
Best Local Similarity 96.9%; Pred. No. 3e-59;  
Matches 340; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 1634 cccggaatttatttattttagtagagagaggtttgcatgttgccagcgtgtctc 1693  
DB 351 cccggaatttatttattttagtagagagaggtttgcatgttgccagcgtgtctc 292  
QY 1694 gaactctgactcagatgacatcgtccttgccctccacagtgctgagattacagc 1753  
DB 291 GAACCTGTGACNMCANATGANCATNTGCTTGCCCTCCACAGTGTGGGATTACAGGC 232  
QY 1754 gttagcacacatgccgacctctctctacaccttaacctgtctcttacctcaacct 1813  
DB 231 GTGAGCCACCATGNCGGGCTCTTCTTCACACTTACACCTGTCCTTATCCACATCT 172  
QY 1814 gttttacaccttcaacctcgtctctctctctctctctctctctctctctctctct 1873  
DB 171 GTTTTACACCTTCACTCCCTGTTTCTCTCAATGTTCACTGTTTCTCCATGTCTATG 112  
QY 1874 ctgaccttcttaccatttggttggaagcagctctctctctgctgttttttttttt 1933  
DB 111 CTCGCTTTCTTACATTTGCTTTGGAAGGACATNTTCTCTGCTGTTTGTGTTT 52  
QY 1934 cccagaaatcagatattatttttaataaagaanaaacattcctagaagatg 1984  
DB 51 CCCAGAAATCAGTATTATTATTTAAATAGAAAACATTCCTAGAGATG 1

RESULT 14  
AAA43677/c

ID AAA43677 standard; cDNA: 373 BP.

AC AAA43677;

DT 21-AUG-2000 (first entry)

DE Human secreted expressed sequence tag SEQ ID NO:252.

XX Human: mouse; chicken; rat; secreted expressed sequence tag; SESN;  
XX expressed sequence tag; EST; probe; chemotactic; proliferative;  
XX immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;  
XX thrombolytic; antiinflammatory; cytosolic; antibacterial; antitumoral;  
XX antiviral; antidiabetic; antiasthmatic; vulnary; antiparkinsonian;  
XX antilicer; osteopathic; neuroprotective; nootropic; antipsoriatic;  
XX cerebroprotective; anticonvulsant; antidepressant; gene therapy;  
XX vaccine; autoimmune disorder; multiple sclerosis; allergic condition;  
XX insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;  
XX lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;  
XX central nervous system disorder; Alzheimer's disease; stroke;  
XX Parkinson's disease; Huntington's disease; coagulation disorder;  
XX haemophilia; thrombosis; inflammatory disorder; Crohn's disease;  
XX tumour; infection; depression; psoriasis; ss.

OS Homo sapiens.

PN WO200021991-A1.

XX 20-APR-2000.  
PD 15-OCT-1999; 99MO-US24206.  
PF 15-OCT-1999; 98US-0104436.  
PR 15-OCT-1998; 98US-0104436.  
XX (GEM ) GENETICS INST INC.  
PA Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;  
PI Werberg D, Treacy M, Bowman MR;  
XX MPI: 2000-317938/27.  
DR Isolated polynucleotides, and encoded proteins, comprising secreted  
PT expressed sequence tags (SESTs), useful for treating various disorders  
PT such as autoimmune, infectious, and central nervous system disorders -  
XX Claim 1; Page 267-268; 803pp; English.

CC AAA43426 to AAA45925 represent specifically claimed secreted expressed  
CC sequence tags (SESTs), isolated from human, mouse, chicken and rat  
CC tissue sources. The SESTs can have a range of activities depending on  
CC the tissues they were isolated from. The activities include:  
CC chemotactic; proliferative; immunomodulatory; haematopoietic;  
CC chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory;  
CC cytosolic; antibacterial; antifungal; antiviral; antidiabetic;  
CC antiasthmatic; vulnary; antilicer; osteopathic; neuroprotective;  
CC nootropic; antiparkinsonian; antipsoriatic; cerebroprotective;  
CC anticonvulsant; and antidepressant. The SESTs can be used for gene  
CC therapy and in vaccines. The SESTs are useful as probes for the  
CC identification and isolation of full-length cDNAs and genomic DNA  
CC molecules which correspond to the SESTs. Proteins encoded by the SESTs  
CC are useful in assays for determining biological activity and raising  
CC antibodies. They may be useful for treatment of autoimmune disorders  
CC (multiple sclerosis, insulin dependent diabetes), allergic conditions  
CC (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,  
CC osteoporosis, osteoarthritis, central nervous system disorders  
CC (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation  
CC disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's  
CC disease), tumours, bacterial, fungal or viral infections, depression and  
CC psoriasis. AAA45926 to AAA45931 represent linker variants which are given  
CC in the exemplification of the present invention.

SO Sequence 373 BP; 96 A; 121 C; 74 G; 82 T; 0 other;

Query Match 8.7%; Score 337.2; DB 21; Length 373;  
Best Local Similarity 98.9%; Pred. No. 1.1e-58;  
Matches 350; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 3222 ctgcagacagagatcgaatcgaatcgtttattgttggtggtgctctcagaagaacct 3281  
DB 373 CTCGAGACAGGAGATCGAATGCACTTGTATTTGTTGGGGGCTCTCAGAGAACCT 314  
QY 3282 gttaagagcctttagcatttatttgctgtgaagatcctctggaggtgtgatacat 3341  
DB 313 GTGAAGCCTTTATCACTCTTTATTTGCTGTGAAGTCTCTGGAGGTGGATCAT 254  
QY 3342 ttgaagcaagtgactcaatttgaaggaagctcttgaagaagagcgtgtgagcatctg 3401  
DB 253 TTGAAGCAAGTACTTCACTTGAAGGCAAGTCTCTGAAAGAGCTGAGGATCTGG 194  
QY 3402 cagctacacatcgctgtgtagtgtcttggtggtggtggtcctggcactgctgtgaag 3461  
DB 193 CACCTACCAACGCTGTGAGTGTGTGGGCTGGGGCTGCGCACATGCTGTGTGAAG 134  
QY 3462 gatctggcagggagaccacagcg-cccctactgaacatagatgtcagtgcatttaa 3520  
DB 133 GATCTGCGACGAGCACACACAGCCCTCTACTGATCAACCATGATGCACTGCAATTAA 74  
QY 3521 agccatcgacgtgaggggacactgagatgtctctgagattactcagagaagca 3574

Db 73 AGCCATGACGTGAGGGGCCACTGAGATTGCTCTGAGATTACAAAGAATA 20

RESULT 15

AAS58007/C  
ID AAS58007 standard; cDNA; 279 BP.

XX AAS58007;

DT 13-FEB-2002 (first entry)

DE cDNA #683 encoding portion of a human colon tumour protein.

XX Human; colon tumour protein; colon cancer; gene therapy; cytostatic; ss.

OS Homo sapiens.

PN M0200173027-A2.

PD 04-OCT-2001.

PF 22-MAR-2001; 2001WO-US09246.

PR 24-MAR-2000; 2000US-191597P.

PR 04-MAY-2000; 2000US-202024P.

PR 05-MAY-2000; 2000US-202189P.

PA (CORI-) CORIXA CORP.

PI Meagher MJ, Xu J, King GE;

DR WPI; 2001-611627/70.

XX New colon tumour proteins and related nucleic acid, useful for

PT treatment, prevention, diagnosis and monitoring of cancer -

XX Claim 4; Page 165; 299pp; English.

XX The present invention relates to the isolation of novel cDNA sequences

CC encoding for at least an immunogenic portion of human colon tumour

CC proteins. The sequences of the invention are useful in pharmaceutical

CC compositions and vaccines for the prevention and treatment of cancers

CC such as colon cancer. They are also useful for the diagnosis and

CC monitoring of such cancers. Antibodies to the colon tumour proteins

CC and antigen presenting cells that express polynucleotides encoding

CC colon tumour proteins can be used to inhibit the development of

CC cancers. T-cells that react specifically with colon tumour proteins

CC are useful for removing tumour cells from samples (e.g. blood) and

CC for cancer treatment. The polynucleotide sequences are also useful in

CC gene therapy. AAS57325-AAS58880 represent the cDNA sequences of the

CC invention that encode for portions of human colon tumour proteins.

XX Sequence 279 BP; 105 A; 45 C; 84 G; 45 T; 0 other;

Query Match

Best Local Similarity 6.78; Score 258.6; DB 23; Length 279;  
Matches 273; Conservative 0; Mismatches 4; Indels 2; Gaps 1;

Qy 1681 ccaagctgagctgaactctgactcagatgacatcgtgctgctccacagatgc 1740

Db 279 CCAGGCTGGTCTCGAAGCTTGTGATCGATGATCGATGCTGGCTCCACAGAGC 220

Qy 1741 tggattacagagcgatgacacacatgcccggcctcttcaaccttaacctgtctct 1800

Db 219 TGGGATTACAGGCGTGAGCCGATGCCGCGCTTTTCACCTTACACTGTCTTCT 160

Qy 1301 tatccacacatctgtttcacacct--catccctgtctctccatgltcacactgtct 1858

Db 159 TATCCTCACAATCTGTTTACACCTTTCATCCCTGCTTCTCATGATGTCACACTGTCT 100

Qy 1859 tcccatgttcatagtcgccttcttaccattgtgttgaaagcgctctctctgtct 1918

Db 99 TCCCATGTTTCATAGCTGCTTCTTACCAATTTGGTTGAAGGCGACGCTTCTTGCT 40

Qy 1919 tgatttttggtttcccaagaataatcagatattttt 1957

Db 39 TGTGTTTTGTTTTTCCGAAATACACTATTATTITTT 1

Search completed: June 8, 2002, 00:14:17  
Job time: 11567 sec

E-1.4

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 8, 2002, 00:02:10 ; Search time 61.52 Seconds

(without alignments)  
724.002 Million cell updates/sec

Title: US-09-729-674-2

Perfect score: 2130  
Sequence: 1 MEQGRGAAALALCVACA.....DDEGEVEYVDLLEETS 401

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 11073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: A.Geneseq\_032802.\*

1: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1980.DAT.\*  
2: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1981.DAT.\*  
3: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1982.DAT.\*  
4: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1983.DAT.\*  
5: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1984.DAT.\*  
6: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1985.DAT.\*  
7: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1986.DAT.\*  
8: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1987.DAT.\*  
9: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1988.DAT.\*  
10: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1989.DAT.\*  
11: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1990.DAT.\*  
12: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1991.DAT.\*  
13: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1992.DAT.\*  
14: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1993.DAT.\*  
15: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1994.DAT.\*  
16: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1995.DAT.\*  
17: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1996.DAT.\*  
18: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1997.DAT.\*  
19: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1998.DAT.\*  
20: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1999.DAT.\*  
21: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA2000.DAT.\*  
22: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2128	99.9	401	AAV17219	Human secreted pro
2	2128	99.9	401	AAU38989	Human secreted pro
3	2121	99.6	401	AAU38989	Human secreted pro
4	572.5	26.9	804	AAU38989	Human secreted pro
5	571	26.8	736	AAU38989	Human secreted pro
6	571	26.8	736	AAU38989	Human secreted pro
7	571	26.8	736	AAU38989	Human secreted pro
8	567	26.6	736	AAU38989	Human secreted pro
9	539	25.3	363	AAU38989	Human secreted pro
10	532.5	25.0	542	AAU38989	Human secreted pro
11	532.5	25.0	747	AAU38989	Human secreted pro

SEA ID NO: 2  
Database: A.Geneseq\_032802  
Ac NO: AAV17219

12	530.5	24.9	359	22	AAE04200	Human gene 2 encod
13	439	20.6	592	22	AAE04244	Human gene 2 encod
14	258.5	12.1	527	22	AAU38989	Human protein sequ
15	215	10.1	60	22	AAU38989	Novel human secret
16	215	10.1	60	22	AAU38989	Novel human secret
17	199.5	9.4	105	21	AAU58500	Human polypeptide
18	198	9.3	153	22	ABU1197	Lung cancer associ
19	180	8.5	227	22	ABU16396	Human secreted pro
20	139	6.5	139	21	AAU04095	Novel human diago
21	127	6.0	173	22	AAU04217	Human secreted pro
22	100	4.7	1006	22	ABU71507	Human gene 2 encod
23	99	4.6	659	22	AAU06996	Drosophila melanog
24	96.5	4.5	404	21	AAU91465	Sequence of C3Vs.
25	96.5	4.5	489	21	AAU91609	Human secreted pro
26	96.5	4.5	544	21	AAU99373	Human secreted pro
27	96.5	4.5	544	22	AAU29160	Human PRO1374 (UNO
28	96.5	4.5	544	22	AAU29160	Human PRO polypept
29	96.5	4.5	544	22	AAU29160	Protein of the inv
30	94.5	4.4	496	16	AAU21806	Amino acid sequenc
31	94.5	4.4	496	16	AAU21806	Spliced-deleted in
32	93	4.4	1341	21	AAU48448	Mouse neuronal PAS
33	93	4.4	1358	21	AAU48448	Arabidopsis thalia
34	93	4.4	1381	21	AAU48447	Arabidopsis thalia
35	92.5	4.3	806	19	AAU75911	Arabidopsis thalia
36	92	4.3	548	14	AAU33741	Helicobacter leucy
37	91.5	4.3	309	22	ABU11552	XRL. Homo sapiens
38	91	4.3	428	20	AAU08820	Novel human diago
39	91	4.3	1173	22	AAU93171	Staphylococcus sci
40	90.5	4.2	625	22	AAU96547	Human protein sequ
41	90.5	4.2	2408	22	ABU10631	Putative P. abyss
42	90	4.2	531	19	AAU56311	Novel human diago
43	90	4.2	629	22	ABU18854	Protein disulphide
44	89.5	4.2	561	21	AAU49740	Novel human diago
45	89.5	4.2	659	21	AAU49739	Arabidopsis thalia

## ALIGNMENTS

RESULT 1	
AAV17219	standard; Protein: 401 AA.
ID	AAV17219
AC	AAV17219;
DT	09-AUG-1999 (first entry)
DE	Human secreted protein (clone bd306-7).
KW	Secreted protein; kidney; lung; brain; blood; testis; bone marrow;
KW	nutritional activity; cytokine; cell proliferation; immune stimulation;
KW	hematopoiesis regulation; tissue growth; thrombolytic; gene therapy;
KW	anti-inflammatory; tumour invasion.
OS	Homo sapiens.
FT	
FT	Key
FT	Misc-difference 64
FT	Location/Qualifiers
FT	/note="residue indicated as xaa is unspecified"
PN	WO9926961-A1.
PD	03-JUN-1999.
PF	24-NOV-1998; 98WO-US25149.
PR	23-NOV-1998; 98US-0197886.
PR	26-NOV-1997; 97US-0066804.
PA	(GENETICS INST INC.
PI	Agostino MT, Clark HF, Collins-Racie LA, Evans C;
PI	Fechtel K, Jacobs K, Lavallie ER, McCoy JM, Merberg D;

PI Steininger RJ, Treacy M, Wong GG;  
 XX WPI: 1999-357809/30.  
 DR N-PSDB; AAX60801.  
 XX  
 PT New polynucleotides encoding secreted proteins  
 XX  
 PS Claim 10; Page 111-113; 133pp; English.  
 XX  
 CC The invention relates to secreted proteins (AAV17219-228) encoded by  
 CC polynucleotides obtained from human fetal kidney, adult lung, adult  
 CC kidney, adult brain, adult blood, adult testes, and fetal brain and  
 CC murine adult bone marrow cDNA libraries. The secreted protein nucleic  
 CC acid sequences (X6801-811) correspond to clones b0306-7, g1283-6,  
 CC tk317-3, K213-2x, na316-1, nr93-20, np164-1, pe204-1, ya1-1 and yp-1,  
 CC (all clones are deposited as ATCC 98599). The pns and proteins are  
 CC predicted to have biological activities which would make them suitable  
 CC for treating, preventing or ameliorating medical conditions in humans and  
 CC animals, although no supporting data is given. Suggested activities  
 CC include nutritional activity, cytokine and cell  
 CC proliferation/differentiation activity, immune stimulating (e.g. as  
 CC vaccines) or suppressing activity, hematopoiesis regulating activity,  
 CC tissue growth activity, activin/ inhibin activity,  
 CC chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,  
 CC receptor/ligand activity, anti-inflammatory activity, cachectin/tumour  
 CC invasion suppressor activity, and tumour inhibition activity. The pns are  
 CC also stated to be useful for gene therapy.  
 XX  
 SQ Sequence 401 AA;  
 XX  
 Query Match 99.9%; Score 2128; DB 20; Length 401;  
 Best Local Similarity 100.0%; Pred. No. 4.3e-184;  
 Matches 401; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MEGRRGAALALALVACALRAGRAOYERYSPFRSPDELMPIESARHALDKYSGEMW 60  
 Db 1 meprgrgaalalalvacalragraoerysfrsfpdelmpiesayrhaldkysgchw 60  
 QY 61 AESGXYEISIRLRRLRDSFAFCHRNCSAARPEPAAGLASYPELRLEGGLRRACHLK 120  
 Db 61 aesgxyeislrlrlrdsfafchrnCSAARPEPAAGLASYPELRLEGGLRRACHLK 120  
 QY 121 RCRKGALAFRQSPSREVLADFORREPKYKLOFAVYFRKANNLPKALAAHPTFLKHPDEM 180  
 Db 121 rckgaldafRQSPSREVLADfgrrePKYKlOFAVYfRkannlpkaiaahtflkhpddem 180  
 QY 181 MKRMMAYKSLPGAEDYIKDLETSYSLEFTRAVAYNGENMRTSITDMELALPDFEKAF 240  
 Db 181 mkrmaykyslpgaedYIKDLETSysleFtravayngemrTStidmElalpDfEkaF 240  
 QY 241 YECLAACEGSEKEIDKFDFYISADHYVEVELECKIOCEENTPIVIGYVPEKEFYATMYHY 300  
 Db 241 yecLaacegseKEIDKFDFYISADHYVEVELECKIOCEentPIVIGYVPEKEFYATMYHY 300  
 QY 301 LOFAYYTLNDKNAAPCAVSYLLFDONDKVMQNLVYQYHRDWMGSDENHFORPEAVQ 360  
 Db 301 lOFAyYtLNDKnaAPCAVsyLLfDONdKVMQNLVYqYHRdWmGSdENHfORPeAVQ 360  
 QY 361 FFNVYTLQKELYDFAKENIMDDGGEVVEYVDLLLEETS 401  
 Db 361 ffnvYtLqkELyDfAKENIMDDGGEVVEYVDLLleETS 401  
 RESULT 2  
 AAU38989  
 ID AAU38989 standard; Protein; 401 AA.  
 XX  
 AC AAU38989;  
 XX  
 DT 16-JAN-2002 (first entry)  
 XX  
 DE Human secreted protein b0306\_7.

XX  
 KW Human; secreted protein; antiinflammatory; immunosuppressive;  
 KW nootropic; neuroprotective; antiarthritic; antimicrobial; vulnary;  
 KW cytosatic; antidiabetic; virucide; antiinfectivity; anticonvulsant;  
 KW vasotropic; antiparkinsonian; immunostimulant; dermatological;  
 KW antirheumatic; antitumor; antulcer; osteopathic; tranquiliser;  
 KW cerebroprotective; cytokine; cell proliferation; cell differentiation;  
 KW immune deficiency; severe combined immunodeficiency; SCID; tumour;  
 KW autoimmune disorder; multiple sclerosis; rheumatoid arthritis;  
 KW graft-versus-host disease; myeloid deficiency; wound healing; ulcer;  
 KW periodontal disease; osteoporosis; osteoarthritis; Alzheimer's disease;  
 KW Parkinson's disease; Huntington's disease; infection; cardiac disease;  
 KW stroke; sepsis; inflammatory bowel disease; contraceptive; immunogen;  
 KW food supplement; vaccine.  
 KW  
 OS Homo sapiens.  
 XX  
 XX WO200175068-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 XX 22-MAR-2001; 2001WO-US09369.  
 XX  
 PR 30-MAR-2000; 2000US-0539330.  
 PR 04-DEC-2000; 2000US-0729674.  
 XX  
 PA (GEMV ) GENETICS INST INC.  
 XX  
 PI Jacobs K, McCoy JM, Lavallie E, Collins-racie LA, Evans C;  
 PI Treacy M, Agostino MJ, Steininger RJ, Spaulding V, Wong GG;  
 PI Clark H, Fechtel K, Merberg D;  
 XX  
 DR WPI: 2001-639363/73.  
 N-PSDB; AAS59207.  
 XX  
 PT Secreted human proteins, useful as vaccine for treating various  
 PT diseases such as autoimmune disorders (e.g. multiple sclerosis), and  
 PT nervous system disorders (e.g. stroke) .  
 XX  
 PS Claim 2; Page 457-459; 619pp; English.  
 XX  
 CC The invention relates to novel human secreted proteins, the nucleic  
 CC acids encoding them. The protein may exhibit cytokine, cell proliferation  
 CC or cell differentiation activity or may induce production of other  
 CC cytokines in certain cell populations and may exhibit immune stimulating  
 CC or immune suppressing activity, which is useful for the treatment of  
 CC various immune deficiencies and disorders e.g. severe combined  
 CC immunodeficiency (SCID), autoimmune disorders e.g. multiple sclerosis,  
 CC systemic lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary  
 CC inflammation. The proteins are also useful in the treatment of diseases  
 CC and disorders including tissue, skin and organ transplantation and in  
 CC graft-versus-host diseases (GVHD), in the induction of tumour immunity,  
 CC myeloid or lymphoid cell deficiencies, wound healing and tissue repair,  
 CC in the treatment of burns, incisions and ulcers; as well as in treatment  
 CC of periodontal disease, osteoporosis or osteoarthritis, mediated by  
 CC inflammatory processes, diseases of the peripheral nervous system,  
 CC Alzheimer's, Parkinson's disease, Huntington's disease,  
 CC amyloidrophic lateral sclerosis, and Shy-Drager syndrome, infections,  
 CC infection of cardiac and central nervous system vessel e.g. stroke,  
 CC sepsis, inflammatory bowel disease, ulcers, bone regeneration. The  
 CC protein, having activin- or inhibin-related activities is useful as a  
 CC contraceptive based on the ability of inhibins to decrease fertility in  
 CC female mammals and decrease spermatogenesis in male mammals. The  
 CC proteins and nucleic acids are also useful as food supplements. The  
 CC present sequence represents a secreted protein of the invention.  
 XX  
 SQ Sequence 401 AA;  
 XX  
 Query Match 99.9%; Score 2128; DB 22; Length 401;  
 Best Local Similarity 100.0%; Pred. No. 4.3e-184;  
 Matches 401; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



PF 28-JUL-2000; 2000EP-011612.6.  
 XX  
 PR 29-JUL-1999; 99JP-024803.6.  
 PR 27-AUG-1999; 99JP-030025.3.  
 PR 11-JAN-2000; 2000JP-011877.6.  
 PR 02-MAY-2000; 2000JP-018376.7.  
 PR 09-JUN-2000; 2000JP-024189.9.  
 XX  
 PA (HELI-) HELIX RES INST.  
 XX  
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX  
 DR WPI; 2001-318749/34.  
 XX  
 PT Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the  
 PT full-length cDNAs -  
 PT  
 XX  
 PS Claim 8; SEQ ID 15160; 2537PP + CD ROM; English.

CC The present invention describes primer sets for synthesizing 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dT primer; and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesizing polynucleotides  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.

	Query Match	26.9%	Score 572.5	DB 22	Length 804
	Best Local Similarity	35.3%	Pred. No. 6.7e-43		
	Matches 145	Conservative 65	Mismatches 158	Indels 43	Gaps 12
QY	6	RCGAALLLALLVCACALRRGAAQERYSPFRSDLEMLPESRYRHAAIDKVGSEHMAESXG	65		
Db	4	ralkliltlllavaa--asqaeveseagwgmvtpll----faegtaayarqgdpyyvl	56		
QY	66	YLELSRLRLRLDSEAFCHRNCSA-----APOPEPAGLASYPELRLFCGLLR	115		
Db	57	smeralstreraalralrlrcrtqcaadfpweldpdkspspagagsgaalrldsfqgliltr	116		
QY	116	AHCLKRCCKGGLPAFRQSPSPREVLADFQRREPKYKLFQAFYFKANNLPRAIAAHTFLFKH	175		
Db	117	aacrlrc--lppnaahsl--seemelefrkspgylqyayfinklekavaaahfftygn	173		
QY	176	PDDEMKKRNMAVYKSLPCGAEDY-IDLETKSTESLFITAVRAYNGENMTSTTDMELAP	234		
Db	174	pehmemqgnldygtmsgvkheadfkldetqphmgcfrlgyvrylseeqgeavaphltaeaql	233		
QY	235	DFFKFYKGLACAESSESLKDKF-----DFSLTADHYEVLECKICEENL-TPVIGG	287		
Db	234	eyfayaecealacegpydydgyngleyanadlfqaltldhyldylnckqncvtealashpsre	293		
QY	288	YPERKFAVTMYHLOFAFYAKLINDLEKNAAPCAVSYLLFDONDKVMQOONLVYVOYHNDTGL	347		

[illegible]

RESULT	5
AAB93142	
ID	AAB93142 standard; Protein; 736 AA

DT 26-JUN-2001 (first entry)

DE	Human protein sequence	SEQ ID NO:12045.
DE		

KW Human; primer; detection; diagnosis; antisense therapy; gene therapy

OS Homo sapiens.

PN EP1074617-A2.

PD 07-FEB-2001.

28-JUL-2000; 2000EP-0116126.

PR 29-JUL-1999; 99JP-0248036.

PR 11-JAN-2000; 2000JP-0118776.

PR 09-JUN-2000; 2000JP-0241899.

PA (HELI-) HELIX RES INST.

PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J,

XX

XX

PT Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
PT full-length cDNAs -

PS Claim 8; SEQ ID 12045; 2537pp + CD ROM; English

The present invention describes primer sets for synthesising 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesising polynucleotides particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95993 represent human amino acid sequences; and AAH31629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

Sequence	736 AA;
SQ	



Query Match 26.8%; Score 571; DB 22; Length 736;  
 Best Local Similarity 34.4%; Pred. No. 8.1e-43;  
 Matches 145; Conservative 68; Mismatches 168; Indels 40; Gaps 11;

```

QY 6 RGAATLALCVACALRAGRAQYERSFRSPDEMLPLESAVRNALDKSGEHWASXG 65
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
D 4 ralklltlllavvaa--aagaeveasegwmvprdl1-----faegtaayardwpgvvl 56
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 66 YLEISLRHLRLDSEAFCHRNCSA-----APQPPAAGLASYPELRPGGLRR 115
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
D 57 smeralrsraalralrlrcrtqcaadfpwelddwspspagaalrldsfqglrr 116
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 116 AHCLKCKOGSLPAPRQSQPREVLADPQREPRKFLQFAVEKNNLPKATAAHTFLKH 175
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
D 117 aaelrrc-lyppaahsl--seemelefrkspynylqavayfinklekavaaahltfvgn 173
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 176 PDDEMKRNNAAYKSLPGADY- IKDLETKSYSLFIRAVRAYNGENMRTSIDMELAP 234
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
D 174 pehmengnldyyqtmagvkeadfkletqphmgefflgyrllyseegpgavphleaalq 233
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 235 DFKATYECLACEGSEIKDFK-----DFYLSADHYVEVLECKIOCEENL-TPYIGG 287
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
D 234 eyfwayeacralcepgdydgynyleynadlfqaltchylqynckqncvtelashpsre 293
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 288 YPEKEFVATWYHYLOFAYYKLNLDKNAAPCAVSYLFDQDKYMOONLVYQYHRTWGL 347
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
D 294 kpfedflpsihnylqfayynlgnytqavecakyllffpndevmnqlayaaam-----l 348
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 348 SDEHFQ---PRPAVOFFNVTTLQKELYDFAKE---NIMDDGEVEVEVDLLELEET 400
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
D 349 geehttsigpresakeyrgsllekellffaydvfglfpvdpswpreeviprllqekq 408
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 401 S 401
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
D 409 s 409
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 6  
 AAB93215  
 ID AAB93215 standard; Protein; 736 AA.  
 AC AAB93215;  
 XX  
 DT 26-JUN-2001 (first entry)  
 XX  
 DE Human protein sequence SEQ ID NO:12194.  
 XX  
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EPI074617-A2.  
 XX  
 PD 07-FEB-2001.  
 XX  
 PF 28-JUL-2000; 2000EP-0116126.  
 XX  
 PR 29-JUL-1999; 99UP-0248036.  
 XX  
 PR 27-AUG-1999; 99UP-0300253.  
 XX  
 PR 11-JAN-2000; 2000JP-0118776.  
 XX  
 PR 02-MAY-2000; 2000JP-0183767.  
 XX  
 PR 09-JUN-2000; 2000JP-0241899.  
 XX  
 PA (HELI-) HELIX RES INST.  
 XX  
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 XX  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX  
 DR WPI; 2001-318749/34.  
 XX  
 PT Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection

PT and/or diagnosis of the abnormality of the proteins encoded by the  
 PT full-length cDNAs -  
 XX  
 PS Claim 8; SEQ ID 12194, 2537pp + CD ROM; English.  
 XX

The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-AT primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.

Sequence 736 AA;

Query Match 26.8%; Score 571; DB 22; Length 736;  
 Best Local Similarity 34.4%; Pred. No. 8.1e-43;  
 Matches 145; Conservative 68; Mismatches 168; Indels 40; Gaps 11;

```

QY 6 RGAATLALCVACALRAGRAQYERSFRSPDEMLPLESAVRNALDKSGEHWASXG 65
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
D 4 ralklltlllavvaa--aagaeveasegwmvprdl1-----faegtaayardwpgvvl 56
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 66 YLEISLRHLRLDSEAFCHRNCSA-----APQPPAAGLASYPELRPGGLRR 115
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
D 57 smeralrsraalralrlrcrtqcaadfpwelddwspspagaalrldsfqglrr 116
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 116 AHCLKCKOGSLPAPRQSQPREVLADPQREPRKFLQFAVEKNNLPKATAAHTFLKH 175
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
D 117 aaelrrc-lyppaahsl--seemelefrkspynylqavayfinklekavaaahltfvgn 173
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 176 PDDEMKRNNAAYKSLPGADY- IKDLETKSYSLFIRAVRAYNGENMRTSIDMELAP 234
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
D 174 pehmengnldyyqtmagvkeadfkletqphmgefflgyrllyseegpgavphleaalq 233
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 235 DFKATYECLACEGSEIKDFK-----DFYLSADHYVEVLECKIOCEENL-TPYIGG 287
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
D 234 eyfwayeacralcepgdydgynyleynadlfqaltchylqynckqncvtelashpsre 293
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 288 YPEKEFVATWYHYLOFAYYKLNLDKNAAPCAVSYLFDQDKYMOONLVYQYHRTWGL 347
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
D 294 kpfedflpsihnylqfayynlgnytqavecakyllffpndevmnqlayaaam-----l 348
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 348 SDEHFQ---PRPAVOFFNVTTLQKELYDFAKE---NIMDDGEVEVEVDLLELEET 400
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
D 349 geehttsigpresakeyrgsllekellffaydvfglfpvdpswpreeviprllqekq 408
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 401 S 401
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
D 409 s 409
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 7  
 AAB88373  
 ID AAB88373 standard; Protein; 736 AA.  
 AC AAB88373;  
 XX  
 DT 26-JUN-2001 (first entry)  
 XX  
 DE Human protein sequence SEQ ID NO:12194.  
 XX  
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EPI074617-A2.  
 XX  
 PD 07-FEB-2001.  
 XX  
 PF 28-JUL-2000; 2000EP-0116126.  
 XX  
 PR 29-JUL-1999; 99UP-0248036.  
 XX  
 PR 27-AUG-1999; 99UP-0300253.  
 XX  
 PR 11-JAN-2000; 2000JP-0118776.  
 XX  
 PR 02-MAY-2000; 2000JP-0183767.  
 XX  
 PR 09-JUN-2000; 2000JP-0241899.  
 XX  
 PA (HELI-) HELIX RES INST.  
 XX  
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 XX  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX  
 DR WPI; 2001-318749/34.  
 XX  
 PT Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection

```

XX DT 23-MAY-2001 (first entry)
XX XX Human membrane or secretory protein clone PSEC0109.
DE DE Human; secretory protein; membrane protein; vaccine; gene therapy;
KW KW rheumatoid arthritis; diabetes.
XX OS Homo sapiens.
XX PN EPI067182-A2.
XX PD 10-JAN-2001.
XX PE 07-JUL-2000; 2000EP-0114090.
XX PR 08-JUL-1999; 99JP-0194179.
XX PR 11-JAN-2000; 2000JP-0118775.
XX PR 02-MAY-2000; 2000JP-0183766.
XX PA (HELI-) HELIX RES INST.
XX PI Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
XX DR WPI: 2001-093989/11.
XX DR N-PSDB: AAF93800.
XX PT Nucleic acids encoding secretory proteins/membrane proteins, useful in
XX PT gene therapy or as candidate target molecules in drug development -
XX PS Claim 1; SEQ ID 114; 609pp + CD ROM; English.
XX CC This invention relates to nucleic acid sequences AAF93744 - AAF93916
XX CC which encode human secretory or membrane proteins represented by
XX CC AAB88317 - AAB88419. Included in the invention are primers
XX CC AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the
XX CC cDNA sequences of the invention. The invention also includes methods for
XX CC the production of antibodies directed against the proteins, and cDNA
XX CC sequences, which can be used in vaccines. The polynucleotide sequences
XX CC can be used in gene therapy. The polynucleotide sequences and the
XX CC proteins they encode may be used in the prevention, treatment and
XX CC diagnosis of diseases associated with inappropriate secretory
XX CC protein/membrane protein expression. The nucleic acids and complementary
XX CC sequences may also be used as DNA probes in diagnostic assays
XX CC (e.g. polymerase chain reactions (PCR)) to detect and quantitate the
XX CC presence of similar nucleic acid sequences in samples. They may also be
XX CC used to study the expression and function of secretory proteins/membrane
XX CC polypeptides and their role in metabolism. The polypeptides may be used
XX CC as antigens in the production of antibodies against them and in assays to
XX CC identify modulators (agonists and antagonists) of expression and
XX CC activity. The antibodies and antagonists may also be used as therapeutic
XX CC agents to down regulate expression and activity. The antibodies may also
XX CC be used as diagnostic agents for detecting the presence of the
XX CC polypeptides in samples (e.g. by enzyme linked immunosorbant assay
XX CC (ELISA)). Examples of diseases which may be treated include rheumatoid
XX CC arthritis and diabetes.
XX SQ Sequence 736 AA;

```

Query Match 26.8%; Score 571; DB 22; Length 736;  
 Best Local Similarity 34.4%; Pred. No. 8.1e-42;  
 Matches 145; Conservative 66; Mismatches 168; Indels 40; Gaps 11;

```

OY 6 RGAATLALLCVACALAGRAQYERYSFRSPRDELMPLESAYRHALDKYSGEHWAESXG 65
DB 4 ralklltlllavvaa--asqevsesagwgmvtpdll-----faegtaayargdwpyvl 56
OY 66 YLEISLRLHLRLDSEAFCHNCSA-----APOPEPAAGLASYPELIRLFGGLIRR 115
DB 57 smeraltsraalralrlrctqcaadfpwelqpdwspspagaasaaalrldlsffgllrr 116
OY 116 AHCLKRKQGLPAPRQSGPSREVLADFGRRRPFYFLQFAVYKANNLRKALNAHTFLK 175

```

```

DB 117 aacitrc-1gppaahsl--seemelefrkspnylqvaayfikinkakavaaahftfygn 173
OY 176 PDDENKRNMAAYKSLUPGAEDY-IRKLETKSYESLFIKRAVYNGENMFTSTIDNELAP 234
DB 174 pehmengnldyqlumsgykheadfkldetqpmgetrigrvrllyseegpqgavphleaalq 233
OY 235 DEFKAFYECLAAEGSREIKDFK-----DFYLSADHVEVELECKIOCEENL-TPVIGG 287
DB 234 eyfvaeeccralcepgydygynyleynadlfqaltldhyigvlnckqncvtelaashpre 293
OY 268 YPEKEFVATMYHYLOFAYYKLNLDLKNAPCAVSYLLFPQDNKVMQONLVYYGHHDTWGL 347
DB 294 kpfedflpslhylnyldfayynlgnylqavecacktyllffpndevmngnlayaam-----1 348
OY 348 SDEHFQ---PREAVQFVNTLQKELYDFAKE---NIMDDGECEVVEYVDLLEDET 400
DB 349 geehtsrsgpresakeyqrsilekellffaydvrglfpvdpsvtrpeevlprkrlqexqk 408
OY 401 S 401
DB 409 s 409

```

RESULT 8  
 AAB36392  
 ID AAB36392 standard; Protein; 736 AA.  
 XX AC AAB36392;  
 XX DT 27-FEB-2001 (first entry)  
 XX XX Human tumour suppressor Grosl-S protein SEQ ID NO:4.  
 DE DE Tumour suppressor; Grosl-L; Grosl-S; cell proliferation; regulation;  
 KW KW cancer; cytostatic; gene therapy.  
 XX OS Homo sapiens.  
 XX PN WO200065047-A1.  
 XX PD 02-NOV-2000.  
 XX PF 26-APR-2000; 2000WO-JP02731.  
 XX PR 26-APR-1999; 99JP-0118806.  
 XX PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.  
 XX PI Wadhwa R, Sugihara T, Yoshida A;  
 XX DR WPI: 2000-687340/67.  
 XX DR N-PSDB: AAC64725.  
 XX PT Full-length tumor suppressor genes encoding Grosl-L, Grosl-S and mouse  
 XX PT homologs participating in regulation of cell proliferation, useful in  
 XX PT development of preventives and remedies of cancer -  
 XX PS Claim 1; Page 78-83; 114pp; Japanese.  
 XX CC The present sequence represents the human tumour suppressor designated  
 XX CC Grosl-S. Grosl-L and Grosl-S have cytostatic activity and can be used  
 XX CC in gene therapy. Grosl-L and Grosl-S genes are useful in the development  
 XX CC of drugs used to treat and prevent cancer.

Query Match 26.6%; Score 567; DB 21; Length 736;  
 Best Local Similarity 34.2%; Pred. No. 1.9e-42;  
 Matches 144; Conservative 68; Mismatches 169; Indels 40; Gaps 11;

```

OY 6 RGAATLALLCVACALAGRAQYERYSFRSPRDELMPLESAYRHALDKYSGEHWAESXG 65

```

```
Db      4 ralklltlllavvaa--asgaeveseagwgmvtprdl1-----faegtaayargdwpqvv1 56
OY      66 YLEISLRHLRLDSEAFCHRNCSA-----APQPPAAGLASYPELRFGGLRR 115
       57 smeralrstraalralrlrctgcgaadfpwldpdpwspspgaagaldrlsf99llrr 116
OY      116 AHLCKRCQGLPAPFROSQSPREVLADPQREPKFLQFAFKANNLPKATAAHTFLKH 175
       117 aaelrrc-lyppaahs1--seemelefrksypynglyvayfklnklekavaaahltffvgn 173
OY      176 PDDEMKRNNAAYKSLPGAEDY-1KDETKSYESLFRAYRAYNGENWRTSITDMELAP 234
       174 pehmenqunldyyqtsmgyvkeadfkldetqphmgfeirlgyrlyseegpqaavphleaalq 233
OY      235 DFFKAYECLAAEGSREIKDFK-----DFYLSIADHYVEVECKIOCEENL-TPVIGG 287
       234 eyfwayeecalcegydygynyleynadlfqalcchylqvinckqncvtelashpsre 293
OY      288 YPVEKTVATMYHLYQFAFYKLNLDKNAAPCAVSYLLEFDONDKYMOONLVYQYHRDTWGL 347
       294 kpfedflpsihnylyqfayynignytqagacakyllffpndevmngnlayaam-----1 348
OY      348 SDEHFO--PRPEAVQFFNVYTLQKELYDPAKE---NIMDDDEGEVEVEYVDLLELEET 400
       349 geehtsrsgpresakeyqrtsllekel1ffaydvfgipfvdpdswpreevlpkrldqekq 408
OY      401 S 401
Db      409 s 409

RESULT 9
AAB36391
ID      AAB36391 standard; Protein; 363 AA.
XX
AC      AAB36391;
XX
DT      27-FEB-2001 (first entry)
XX
DE      Human tumour suppressor Gros1-L protein SEQ ID NO:2.
XX
KW      Tumour suppressor; Gros1-L; Gros1-S; cell proliferation; regulation;
KW      cancer; cytostatic; gene therapy.
XX
OS      Homo sapiens.
XX
PN      WO200065047-A1.
XX
PD      02-NOV-2000.
XX
PF      26-APR-2000; 2000MO-JP02731.
XX
PR      26-APR-1999; 99JP-0118806.
XX
PA      (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX
PI      Wadhwa R, Sugihara T, Yoshida A;
XX
DR      WPI: 2000-687340/67.
DR      N-PSDB: AAC64724.
XX
PT      Full-length tumor suppressor genes encoding Gros1-L, Gros1-S and mouse
PT      homologs participating in regulation of cell proliferation, useful in
PT      development of preventives and remedies of cancer
XX
XX      Claim 1; Page 67-69; 114pp; Japanese.
XX
CC      The present sequence represents the human tumour suppressor designated
CC      Gros1-L. Gros1-L and Gros1-S have cytostatic activity and can be used
CC      in gene therapy. Gros1-L and Gros1-S genes are useful in the development
CC      of drugs used to treat and prevent cancer.
XX
```

```
SO      Sequence      363 AA;
Query Match      25.3%; Score 539; DB 21; Length 363;
Best Local Similarity 35.4%; Pred. No. 2,4e-40;
Matches 132; Conservative 59; Mismatches 146; Indels 36; Gaps 10;
OY      6 RGAALLALICVACALRAGRAQYERSFRSPDELMPLSAVRHLDKXSGEHMAESXG 65
       4 ralklltlllavvaa--asgaeveseagwgmvtprdl1-----faegtaayargdwpqvv1 56
Db
OY      66 YLEISLRHLRLDSEAFCHRNCSA-----APQPPAAGLASYPELRFGGLRR 115
       57 smeralrstraalralrlrctgcgaadfpwldpdpwspspgaagaldrlsf99llrr 116
OY      116 AHLCKRCQGLPAPFROSQSPREVLADPQREPKFLQFAFKANNLPKATAAHTFLKH 175
       117 aaelrrc-lyppaahs1--seemelefrksypynglyvayfklnklekavaaahltffvgn 173
OY      176 PDDEMKRNNAAYKSLPGAEDY-1KDETKSYESLFRAYRAYNGENWRTSITDMELAP 234
       174 pehmenqunldyyqtsmgyvkeadfkldetqphmgfeirlgyrlyseegpqaavphleaalq 233
OY      235 DFFKAYECLAAEGSREIKDFK-----DFYLSIADHYVEVECKIOCEENL-TPVIGG 287
       234 eyfwayeecalcegydygynyleynadlfqalcchylqvinckqncvtelashpsre 293
OY      288 YPVEKTVATMYHLYQFAFYKLNLDKNAAPCAVSYLLEFDONDKYMOONLVYQYHRDTWGL 347
       294 kpfedflpsihnylyqfayynignytqagacakyllffpndevmngnlayaam-----1 348
OY      348 SDEHFO--PRPE 357
       349 geehtsrsgpreq 361
Db
```

```
RESULT 10
AAB36394
ID      AAB36394 standard; Protein; 542 AA.
XX
AC      AAB36394;
XX
DT      27-FEB-2001 (first entry)
XX
DE      Mouse tumour suppressor Gros1-S protein SEQ ID NO:8.
XX
KW      Tumour suppressor; Gros1-L; Gros1-S; cell proliferation; regulation;
KW      cancer; cytostatic; gene therapy.
XX
OS      Mus musculus.
XX
PN      WO200065047-A1.
XX
PD      02-NOV-2000.
XX
PF      26-APR-2000; 2000MO-JP02731.
XX
PR      26-APR-1999; 99JP-0118806.
XX
PA      (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX
PI      Wadhwa R, Sugihara T, Yoshida A;
XX
DR      WPI: 2000-687340/67.
DR      N-PSDB: AAC64727.
XX
PT      Full-length tumor suppressor genes encoding Gros1-L, Gros1-S and mouse
PT      homologs participating in regulation of cell proliferation, useful in
PT      development of preventives and remedies of cancer
XX
XX      Claim 1; Page 104-107; 114pp; Japanese.
XX
CC      The present sequence represents the mouse tumour suppressor designated
```

CC Gros1-S, Gros1-L and Gros1-S have cytoskeletal activity and can be used  
 CC in gene therapy. Gros1-L and Gros1-S genes are useful in the development  
 CC of drugs used to treat and prevent cancer.  
 XX  
 XX Sequence 542 AA;

Query Match 25.0%; Score 532.5; DB 21; Length 542;  
 Best Local Similarity 35.4%; Pred. No. 1.6e-39;  
 Matches 135; Conservative 57; Mismatches 154; Indels 35; Gaps 11;

QY 48 YRHADKXSGEHWASXGYLEISLRHLRLDSEAFCHRNCSA----APQ----PEPA-- 97  
 Db 39 yaegtaaystrdwpjvnlmeraltrsaalralrlrcrtcatelpwapddldlppdpsls 98  
 QY 98 --AGLASYPELRLEFGGLRRACLRCKQGLPAFRQSQPSREVLADFORREPYKELQPAY 155  
 Db 99 qdpgaaahldlrrffgavltracrlrc-1gppsahl--seeldlefkspynlyqay 155  
 QY 156 FKANNLPKALAAHTFLKHPDDEMKRNMAVYKSLPGAEDY-1KDETKYESLFIKAV 214  
 Db 156 fkinlekavaaahltffvgnpehmemrgnlidyqtmgsvkheadfrdleakphmhefrlgy 215  
 QY 215 RAYNGENRTSTDMELALDPFKAFYECCLACEGSREIKDPK-----DYLSTADHYV 268  
 Db 216 rlyseekpqeavphnealqeyfvadeecralcegydydgyndlydsadlfgaltidhyv 275  
 QY 269 EVLECKIOCEENL-TPVIGGYRVEKFVATMYHYLOFAYYKINDLKNAAPCAVSYLLFQON 327  
 Db 276 qvlnckqncvelashpsrekrfedflpsbnylyqfayngnyqalecktyllffpn 335  
 QY 328 DKVMQONLVYQYHNDTWGLSDH---FQRPPEAVQFPNVTTLQEL---YDPAKENIM 380  
 Db 336 devmhqnlaly----famlgeeeassisprenaeeeyrrpnlllekellffaydlfglpgfv 390  
 QY 381 DDEGEVEYVDLLEETS 401  
 Db 391 dpdswtpeevlprklqekqs 411

RESULT 11  
 AAB36393  
 ID AAB36393 standard; Protein: 747 AA.

XX AAB36393;  
 XX  
 DT 27-FEB-2001 (first entry)  
 XX  
 DE Mouse tumour suppressor Gros1-L protein SEQ ID NO:6.  
 XX  
 KW Tumour suppressor; Gros1-L; Gros1-S; cell proliferation; regulation;  
 KW cancer; cytoskeletal; gene therapy.  
 XX  
 OS Mus musculus.  
 XX  
 PN NO200065047-A1.  
 XX  
 PD 02-NOV-2000.  
 XX  
 PF 26-APR-2000; 2000MO-JF02721.  
 XX  
 PR 26-APR-1999; 99JP-0118806.  
 XX  
 PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.  
 XX  
 PI Madhwa R, Sugihara T, Yoshida A;  
 XX  
 DR WPI: 2000-687340/67.  
 DR N-PSDB; AAC64726.  
 XX  
 PT Full-length tumor suppressor genes encoding Gros1-L, Gros1-S and mouse  
 PT homologs participating in regulation of cell proliferation, useful in  
 PT development of preventives and remedies of cancer

XX  
 PS Claim 1; Page 91-97; 114pp; Japanese.  
 XX  
 CC The present sequence represents the mouse tumour suppressor designated  
 CC Gros1-L, Gros1-L and Gros1-S have cytoskeletal activity and can be used  
 CC in gene therapy. Gros1-L and Gros1-S genes are useful in the development  
 CC of drugs used to treat and prevent cancer.  
 XX  
 XX Sequence 747 AA;

Query Match 25.0%; Score 532.5; DB 21; Length 747;  
 Best Local Similarity 35.4%; Pred. No. 2.5e-39;  
 Matches 135; Conservative 57; Mismatches 154; Indels 35; Gaps 11;

QY 48 YRHADKXSGEHWASXGYLEISLRHLRLDSEAFCHRNCSA----APQ----PEPA-- 97  
 Db 39 yaegtaaystrdwpjvnlmeraltrsaalralrlrcrtcatelpwapddldlppdpsls 98  
 QY 98 --AGLASYPELRLEFGGLRRACLRCKQGLPAFRQSQPSREVLADFORREPYKELQPAY 155  
 Db 99 qdpgaaahldlrrffgavltracrlrc-1gppsahl--seeldlefkspynlyqay 155  
 QY 156 FKANNLPKALAAHTFLKHPDDEMKRNMAVYKSLPGAEDY-1KDETKYESLFIKAV 214  
 Db 156 fkinlekavaaahltffvgnpehmemrgnlidyqtmgsvkheadfrdleakphmhefrlgy 215  
 QY 215 RAYNGENRTSTDMELALDPFKAFYECCLACEGSREIKDPK-----DYLSTADHYV 268  
 Db 216 rlyseekpqeavphnealqeyfvadeecralcegydydgyndlydsadlfgaltidhyv 275  
 QY 269 EVLECKIOCEENL-TPVIGGYRVEKFVATMYHYLOFAYYKINDLKNAAPCAVSYLLFQON 327  
 Db 276 qvlnckqncvelashpsrekrfedflpsbnylyqfayngnyqalecktyllffpn 335  
 QY 328 DKVMQONLVYQYHNDTWGLSDH---FQRPPEAVQFPNVTTLQEL---YDPAKENIM 380  
 Db 336 devmhqnlaly----famlgeeeassisprenaeeeyrrpnlllekellffaydlfglpgfv 390  
 QY 381 DDEGEVEYVDLLEETS 401  
 Db 391 dpdswtpeevlprklqekqs 411

RESULT 12  
 AAE04200  
 ID AAE04200 standard; Protein: 359 AA.

XX AAE04200;  
 XX  
 DT 09-AUG-2001 (first entry)  
 XX  
 DE Human gene 2 encoded secreted protein HUVFY29, SEQ ID NO:54.  
 XX  
 KW Human; secreted protein; proliferative disorder; cancer; tumour;  
 KW foetal abnormality; developmental abnormality; haematopoietic disorder;  
 KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
 KW inflammation; allergy; neurological disorder; Alzheimer's disease;  
 KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;  
 KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;  
 KW cardiovascular disorder; angiotensin disorder; kidney disorder;  
 KW gastrointestinal disorder; pregnancy-related disorder; gene therapy;  
 KW endocrine disorder; infection; wound healing; vulnerability;  
 KW cell culture; chemotaxis; food additive;  
 KW binding partner identification.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key  
 FT Peptide  
 FT Protein  
 FT Location/Qualifiers  
 FT 1..18  
 FT /label= signal\_peptide  
 FT 19..359  
 FT /note= "Mature secreted protein"

CC AAD008488-AAD008529 represent cDNAs corresponding to 18 human secreted  
 CC protein genes, and AAE004199-AAE004239 represent the proteins they encode.  
 CC AAE004240-AAE004297 represent human secreted protein fragments or variants  
 CC The secreted proteins and their genes are useful for preventing,  
 CC treating or ameliorating medical conditions, e.g., by protein or gene  
 CC therapy. Pathological conditions can be diagnosed by determining the  
 CC amount of the new protein in a sample or by determining the presence of  
 CC mutations in the new genes. Specific uses are described for each of the  
 CC 18 genes, based on the tissues in which they are most highly expressed,  
 CC and include developing products for the diagnosis or treatment of  
 CC proliferative disorders, cancer, tumours, foetal and developmental  
 CC abnormalities, haematopoietic disorders, diseases of the immune system,  
 CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,  
 CC allergies, neurological disorders (e.g., Alzheimer's disease,  
 CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,  
 CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,  
 CC cardiovascular disorders, angiogenic disorders, kidney disorders,  
 CC gastrointestinal disorders, pregnancy-related disorders, endocrine  
 CC disorders, and infections. The proteins can also be used to aid wound  
 CC healing and epithelial cell proliferation, to prevent skin ageing due to  
 CC sunburn, to maintain organs before transplantation, for supporting cell  
 CC culture of primary tissues, to regenerate tissues, to identify their  
 CC cognate ligands or binding partners, and in chemotaxis, and can be used  
 CC as a food additive or preservative to modify storage properties.  
 CC Antibodies specific for a protein of the invention can be used in  
 CC alleviating symptoms associated with the disorders mentioned above, and  
 CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked  
 CC immunosorbent assay (ELISA). The present sequence represents a human  
 CC secreted protein of the invention.

```

0Y      6  RGAALLTALCVACMLRAGRAQERKSFSPFRODEMLPESLYRRALDKYSEHNHAEKXG  65
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      4  rakliltlllavvaa--asgaeveseagwmvlpdl)----faegtaayagdgpvyvl  56
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
0Y      66  YLEISLRHLRLDSEAFCHRNCSA-----APOEPAAGLASYPETLRFGLLRR  115
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      57  smeralstraalralrlrcrtgcaadfpwelpbdwpspspagsaaalrldstffgylrr  116
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
0Y      116  AHCLCRKQGLPFRQSOPSEVLADFQRERYKKFLQFPEFANMLPRAIAAHFELLK  175
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      117  aaelrrc-igppaahsl---seemaeftirspyyuqayafinklekavaahhflfygn  173
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
0Y      176  PDDEMKRNKAYKSLPGAEDY- IKDLFETKSYESLFIRAVRAYNENMRTSTTDMELALP  234
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

```

CC MAD004888 MAD005829 represent cDNAs corresponding to 18 human secreted  
CC proteins, genes, and MAD004119 MAD004239 represent the proteins they encode.  
CC AAE004240-AAE004297 represent human secreted protein fragments or variants.  
CC The secreted proteins and their genes are useful for preventing,  
CC treating or ameliorating medical conditions, e.g., by protein or  
CC therapy. Pathological conditions can be diagnosed by determining the  
CC amount of the new protein in a sample or by determining the presence of  
CC mutations in the new genes. Specific uses are described for each of the  
CC 18 genes, based on the tissues in which they are most highly expressed,  
CC and include developing products for the diagnosis or treatment of  
CC proliferative disorders, cancer, tumours, foetal and developmental  
CC abnormalities, haematopoietic disorders, diseases of the immune system,  
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,  
CC allergies, neurological disorders (e.g., Alzheimer's disease,





XX WO200179449-A2.  
 PN  
 XX  
 PD 25-OCT-2001.  
 XX  
 PF 16-APR-2001; 2001WO-US08656.  
 XX  
 PR 18-APR-2000; 2000US-0552929.  
 PR 26-JAN-2001; 2001US-0770160.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Tang YT, Liu C, Drmanac RT;  
 XX  
 DR WPI; 2001-611725/70.  
 XX  
 PT Nucleic acids encoding a range of human polypeptides, useful in genetic  
 PT vaccination, testing and therapy -  
 XX  
 PS Claim 20; Page 686; 765pp; English.  
 XX  
 CC The invention relates to novel human secreted polypeptides. The  
 CC polypeptides and antibodies to the polypeptides are useful for  
 CC determining the presence of or predisposition to a disease associated  
 CC with altered levels of polypeptide. The polypeptides are also useful for  
 CC identifying agents (agonists and antagonists) that bind to them. Cells  
 CC expressing the proteins are useful for identifying a therapeutic agent  
 CC for use in treatment of a pathology related to aberrant expression or  
 CC physiological interactions of the polypeptide. Vectors comprising  
 CC the nucleic acids encoding the polypeptides and cells genetically  
 CC engineered to express them are also useful for producing the proteins.  
 CC The proteins are useful in genetic vaccination, testing and  
 CC therapy, and can be used as nutritional supplements. They may be used to  
 CC increase stem cell proliferation; to regulate haematopoiesis; and in  
 CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;  
 CC immune suppression and/or stimulation; as anti-inflammatory agents; and  
 CC in treatment of leukaemias. AAU29510-AAU3304 represent the amino acid  
 CC sequences of novel human secreted proteins of the invention.  
 XX  
 SO Sequence 60 AA;

Query Match 10.1%; Score 215; DB 22; Length 60;  
 Best Local Similarity 73.2%; Pred. No. 4e-12;  
 Matches 41; Conservative 3; Mismatches 12; Indels 0; Gaps 0;  
 QY 272 ECKIQCEENITPVIGYVPEKFAVMYHYQFAVYKLNLIKNAAPCAVSYLLFDQN 327  
 DB 5 eckihckkkispglrsvenfvdmydytqpaykykindlnadpcavryllfdqn 60

Search completed: June 8, 2002, 02:08:03  
 Job time: 7553 sec

THIS PAGE BLANK (1.5PTO)

p. 1, 2

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 7, 2002, 20:59:40 ; Search time 126.38 Seconds  
(without alignments)  
7523.717 Million cell updates/sec

Title: US-09-729-674-1

Perfect score: 3871

Sequence: 1 ttctctctctctctctctt.....aataaaaaaaaaaaaaa 3871

Scoring table:

IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*

1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PCrUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/Backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	466.4	12.0	618	4	US-09-328-111-215 Sequence 215, App
2	228.8	5.9	9365	4	US-09-608-285A-8 Sequence 8, Appl
3	228.8	5.9	14747	4	US-09-608-285A-42 Sequence 42, Appl
4	228.8	5.9	15977	4	US-09-608-285A-59 Sequence 59, Appl
5	227	5.9	11725	2	US-08-756-506-1 Sequence 1, Appl
6	223.6	5.8	4773	3	US-08-884-324-9 Sequence 9, Appl
7	223.6	5.8	11464	3	US-08-884-324-13 Sequence 13, Appl
8	223.6	5.8	28994	3	US-08-884-324-14 Sequence 14, Appl
9	223	5.8	22481	4	US-08-367-841A-43 Sequence 43, Appl
10	223	5.8	22481	5	PCR-US95-07201-43 Sequence 43, Appl
C 11	222.8	5.8	282	1	US-08-133-629-8 Sequence 8, Appl
12	221.8	5.7	15297	4	US-09-817-180-3 Sequence 3, Appl
13	220.2	5.7	5616	2	US-08-996-306-1 Sequence 1, Appl
14	220.2	5.7	5616	4	US-09-338-907-1 Sequence 1, Appl
15	220.2	5.7	5616	4	US-09-218-207-1 Sequence 1, Appl
16	220.2	5.7	56520	4	US-09-338-907-179 Sequence 179, App
17	220.2	5.7	56520	4	US-09-218-207-179 Sequence 179, App
C 18	219.8	5.7	35060	3	US-08-814-095-7 Sequence 7, Appl
C 19	218.4	5.6	84495	4	US-09-797-906-3 Sequence 3, Appl
C 20	218	5.6	4285	4	US-09-040-774-1 Sequence 1, Appl
21	217.6	5.6	1901	4	US-09-338-907-181 Sequence 181, App
22	217.6	5.6	1901	4	US-09-218-207-181 Sequence 181, App
C 23	217.2	5.6	2061	2	US-08-960-023-11 Sequence 11, Appl
24	217.2	5.6	8453	4	US-09-167-681-45 Sequence 45, Appl
25	217.2	5.6	31571	1	US-08-323-443B-1 Sequence 2, Appl
26	217.2	5.6	33526	3	US-08-658-136-2 Sequence 1, Appl
27	217.2	5.6	53577	3	US-08-658-136-1 Sequence 1, Appl

28	216.8	5.6	72604	4	US-09-268-992-7 Sequence 7, Appl
29	216.6	5.6	35100	1	US-08-306-691B-19 Sequence 19, Appl
C 30	216.6	5.6	35100	5	PCR-US93-06251-19 Sequence 354, App
31	216.2	5.6	631	4	US-09-385-982-354 Sequence 6, Appl
C 32	215.8	5.6	87350	3	US-08-781-891-79 Sequence 79, Appl
C 33	215.4	5.6	2713	2	US-08-916-901-6 Sequence 6, Appl
C 34	215.4	5.6	2713	4	US-09-154-602-6 Sequence 6, Appl
35	215.4	5.6	8174	1	US-07-914-281-5 Sequence 5, Appl
36	215.4	5.6	8174	1	US-08-393-246-5 Sequence 5, Appl
37	215.4	5.6	8174	1	US-08-525-058A-5 Sequence 5, Appl
38	215.4	5.6	8174	2	US-08-696-731-5 Sequence 5, Appl
39	215.4	5.6	8174	4	US-09-042-531-5 Sequence 3, Appl
40	215.4	5.6	8174	5	PCR-US91-00899-3 Sequence 3, Appl
41	215.4	5.6	72928	3	US-09-009-913-1 Sequence 1, Appl
42	214.8	5.5	619	4	US-09-385-982-358 Sequence 358, App
C 43	214.6	5.5	2688	2	US-08-909-965C-1 Sequence 1, Appl
44	214.4	5.5	7720	4	US-09-318-448-5 Sequence 5, Appl
45	214.2	5.5	13158	2	US-08-687-080-105 Sequence 105, App

## ALIGNMENTS

```
RESULT 1
US-09-328-111-215/c
; Sequence 215, Application US/09328111
; Patent NO. 6262333
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carlioli III, Eddie
; APPLICANT: Carlioli, Theodore J.
; APPLICANT: Dertl, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/328,111
; CURRENT FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: US 60/088,801
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 215
; LENGTH: 618
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)..(618)
; OTHER INFORMATION: n = A,T,C or G
US-09-328-111-215

Query Match 12.0%; Score 466.4; DB 4; Length 618;
Best Local Similarity 94.8%; Pred. No. 1.le-99;
Matches 506; Conservative 0; Mismatches 25; Indels 3; Gaps 3;

OY 1076 ccagtaacacaggaacttgggctctcgatgacacacccagacactgaagc 1135
DB 534 CNAAGTCCCNAGNCANTGGGGCTTNGATGAGCANTTCCACCAGNCTTAANC 475
OY 1136 agttcagttcttaagtacac-cactccagaagaagc-tgtatgacttt-gctaagga 1192
DB 474 AGTTCAgTCTTAAGTAAAGTCCACCACTCCAGANGAAGCTTGTATGACTTTGGTAAAGA 415
OY 1193 aatatatgatgatgatgaggaaggtgtgagatatgtgatgactcttgaact 1252
```

```
Db 414 AAATATATGATGATGATGAGGAGAGTGTGGAATATGTGATGACCTCTTGAACT 355
Qy 1253 ggaagagaccagtagaccacagcaacaaagagactcctcttgaggcttaaggaacac 1312
Db 354 GGAGAGACACAGTATGACCCACAGCACAAAGAGACTCTCTTGAGGCTTACGAAACAC 295
Qy 1313 agatctcttgctcttcccaacagccagctgttgatacctaagagcctctcttac 1372
Db 294 AGATCTCTTGTCTTCTTCCCAACAGCCAGGCTGTGTATACCTCAGACCTCTCTTAC 235
Qy 1373 tctcaagtgaaagggagagcccccgtctctctaactgcatgcatcaggggtgagcctg 1432
Db 234 TCTCCAAAGTAAAGGAGAGCCCGCTCTCTTAACCTCAATGCAATCAGGGGTGAGCCTG 175
Qy 1433 ccttccatcttcaacactgcacactcatgttcaacactcttccacactttttt 1492
Db 174 CTTTCCATCTTCACACACTGCACACTCATGTTCACACACTCTTTTCTTCTTTT 115
Qy 1493 gagatgagctcgtctctcttgcccaagctgagatgcaatgcaatgcaatgcaatg 1552
Db 114 GAGATGAGTCTGCTCTCTCTGAGGCTGAGTCAATGCAATGCAATGCAATGCAATGCA 55
Qy 1553 caactccgctctcttggttcaagaatctctgcatcagcctccagatcacc 1606
Db 54 CAACCTCGCCTCTGCTGCTGCAAGCAATCTGCTGCAATCAGCCTCCGAGTAC 1
```

## RESULT 2

US-09-608-285A-8  
Sequence 8, Application US/09608285A  
Patent No. 6335013

GENERAL INFORMATION:

APPLICANT: Ford, John

APPLICANT: Mulero, Julio

APPLICANT: Yeung, George

TITLE OR INVENTION: METHODS AND MATERIALS RELATING TO CD39-LIKE

FILE REFERENCE: 28110/36570

CURRENT APPLICATION NUMBER: US/09/608, 285A

PRIOR FILING DATE: 2000-06-30

PRIOR APPLICATION NUMBER: 09/583, 231

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 09/557, 800

PRIOR FILING DATE: 2000-04-25

PRIOR APPLICATION NUMBER: 09/481, 238

PRIOR FILING DATE: 2000-01-11

PRIOR APPLICATION NUMBER: 09/370, 265

PRIOR FILING DATE: 1999-08-09

PRIOR APPLICATION NUMBER: PCT/US99/16180

PRIOR FILING DATE: 1999-07-16

PRIOR APPLICATION NUMBER: 09/350, 836

PRIOR FILING DATE: 1999-07-09

PRIOR APPLICATION NUMBER: 09/273, 447

PRIOR FILING DATE: 1999-03-19

PRIOR APPLICATION NUMBER: 09/244, 444

PRIOR FILING DATE: 1999-02-04

PRIOR APPLICATION NUMBER: 09/122, 449

PRIOR FILING DATE: 1998-07-24

PRIOR APPLICATION NUMBER: 09/118, 205

PRIOR FILING DATE: 1998-07-16

NUMBER OF SEQ ID NOS: 60

SOFTWARE: Patentln Ver. 2.0

SEQ ID NO 8

LENGTH: 9365

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: misc\_feature

LOCATION: (3409)

OTHER INFORMATION: n = adenine or guanine or cytosine or thymidine

NAME/KEY: misc\_feature

LOCATION: (9214)

OTHER INFORMATION: n = adenine or guanine or cytosine or thymidine  
NAME/KEY: misc\_feature  
LOCATION: (9303)  
OTHER INFORMATION: n = adenine or guanine or cytosine or thymidine  
NAME/KEY: misc\_feature  
LOCATION: (9311)  
OTHER INFORMATION: n = adenine or guanine or cytosine or thymidine  
US-09-608-285A-8

Query Match 5.9%; Score 228.8; DB 4; Length 9365;

Best Local Similarity 75.4%; Pred. No. 1.3e-43; Mismatches 47; Gaps 2;

Matches 261; Conservative 34; Indels 4;

```
Qy 1462 tttcacaccatcttctcacccttctttagagatgagctcgtc-tcttgccagg 1520
Db 7413 tattattattattattattattattttttgagagagctgcttgcttcyygg 7472
Qy 1521 ctgagtgcaatgagcgtctcagctcaactgcaactcgcctcttggttcaagcaat 1580
Db 7473 ctgagtgcaatgagcgtctcagctcaactgcaactcgcctcttggttcaagcaat 1580
Qy 1581 tctgctgcatagcctcccgagctcctgagatagagcagtgccacagcccgctta 1640
Db 7533 tctgctgcatagcctcccgagctcctgagatagagcagtgccacagcccgctta 1640
Qy 1641 a---tttgatattttagtaagagaggttttccatgttgccagcgtgctcgaac 1697
Db 7593 attttttgatttttttagtaagagaggttttccatgttgccagcgtgctcgaac 1697
Qy 1698 tcttgacttcagatgcatcctcgttgccctccacagtgctgagatcaagcgctga 1757
Db 7653 tctgagctcagtgatcaccctccctcgcctcccaagtgctgagatcaagcgctga 1757
Qy 1758 gccacacagcccgctcttctcacccttccacacttaccctgctctcat 1803
Db 7713 gccacacagcccgctcttctcacccttcttctgctcttcttcttcttcttctt 7758
```

## RESULT 3

US-09-608-285A-42  
Sequence 42, Application US/09608285A  
Patent No. 6335013

GENERAL INFORMATION:

APPLICANT: Ford, John

APPLICANT: Mulero, Julio

APPLICANT: Yeung, George

TITLE OR INVENTION: METHODS AND MATERIALS RELATING TO CD39-LIKE

FILE REFERENCE: 28110/36570

CURRENT APPLICATION NUMBER: US/09/608, 285A

PRIOR FILING DATE: 2000-06-30

PRIOR APPLICATION NUMBER: 09/583, 231

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 09/557, 800

PRIOR FILING DATE: 2000-04-25

PRIOR APPLICATION NUMBER: 09/481, 238

PRIOR FILING DATE: 2000-01-11

PRIOR APPLICATION NUMBER: 09/370, 265

PRIOR FILING DATE: 1999-08-09

PRIOR APPLICATION NUMBER: PCT/US99/16180

PRIOR FILING DATE: 1999-07-16

PRIOR APPLICATION NUMBER: 09/350, 836

PRIOR FILING DATE: 1999-07-09

PRIOR APPLICATION NUMBER: 09/273, 447

PRIOR FILING DATE: 1999-03-19

PRIOR APPLICATION NUMBER: 09/244, 444

PRIOR FILING DATE: 1999-02-04

PRIOR APPLICATION NUMBER: 09/122, 449

PRIOR FILING DATE: 1998-07-24

PRIOR APPLICATION NUMBER: 09/118, 205

PRIOR FILING DATE: 1998-07-16

NUMBER OF SEQ ID NOS: 60

```

: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 42
:
: LENGTH: 14747
:
: TYPE: DNA
:
: ORGANISM: Homo sapiens
:
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (13641)
: OTHER INFORMATION: n = adenosine or guanine or cytosine or thymidine
:
: OS-09-608-285A-42

```

Query Match	5.9%;	Score 228.8;	DB 4;	Length 14747;
Best Local Similarity	-75.4%;	Pred. No. 1.6e-43;		
Matches 261;	Conservative 34;	Mismatches 47;	Indels 4;	Gaps 2;

[illegible]

```

1 RESULT 4
2 US-09-608-2854-59
3 : Sequence 59, Application US/096082854
4 : Patent No. 6335013
5 :
6 : GENERAL INFORMATION:
7 :
8 : APPLICANT: Ford, John
9 : APPLICANT: Mulero, Julio
10 : APPLICANT: Yeung, George
11 : TITLE OF INVENTION: METHODS AND MATERIALS
12 : TITLE OF INVENTION: POLYPEPTIDES
13 : FILE REFERENCE: 28110/65570
14 :
15 : CURRENT APPLICATION NUMBER: US/09/608,285A
16 :
17 : CURRENT FILING DATE: 2000-06-30
18 :
19 : PRIOR APPLICATION NUMBER: 09/583,231
20 :
21 : PRIOR FILING DATE: 2000-05-26
22 :
23 : PRIOR APPLICATION NUMBER: 09/557,800
24 :
25 : PRIOR FILING DATE: 2000-04-25
26 :
27 : PRIOR APPLICATION NUMBER: 09/481,238
28 :
29 : PRIOR FILING DATE: 2000-01-11
30 :
31 : PRIOR APPLICATION NUMBER: 09/370,265
32 :
33 : PRIOR FILING DATE: 1999-08-09
34 :
35 : PRIOR APPLICATION NUMBER: FCT/US99/16180
36 :
37 : PRIOR FILING DATE: 1999-07-16
38 :
39 : PRIOR APPLICATION NUMBER: 09/350,836
40 :
41 : PRIOR FILING DATE: 1999-07-09
42 :
43 : PRIOR APPLICATION NUMBER: 09/273,447
44 :
45 : PRIOR FILING DATE: 1999-03-19
46 :
47 : PRIOR APPLICATION NUMBER: 09/244,444
48 :
49 : PRIOR FILING DATE: 1999-02-04
50 :
51 : PRIOR APPLICATION NUMBER: 09/122,449
52 :
53 : PRIOR FILING DATE: 1998-07-24
54 :
55 : PRIOR APPLICATION NUMBER: 09/118,205

```

```

; PRIOR FILING DATE: 1998-07-16
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 59
; LENGTH: 15977

```

```

; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: CD39-L4/L66 Gene Sequence
;

```

```

LOCATION: (245)..(461)
NAME/KEY: CDS
LOCATION: (1454)..(1533)
NAME/KEY: CDS
LOCATION: (2734)..(2877)
NAME/KEY: CDS
LOCATION: (4364)..(4439)
NAME/KEY: CDS
LOCATION: (4679)..(4714)
NAME/KEY: CDS
LOCATION: (5326)..(5414)
NAME/KEY: CDS
LOCATION: (5723)..(5802)
NAME/KEY: CDS
LOCATION: (6751)..(6812)
NAME/KEY: CDS
LOCATION: (7758)..(7859)
NAME/KEY: CDS
LOCATION: (8712)..(8852)
NAME/KEY: CDS
LOCATION: (9831)..(9887)
NAME/KEY: CDS
LOCATION: (11613)..(11728)
NAME/KEY: CDS
LOCATION: (13146)..(13691)
NAME/KEY: CDS
LOCATION: (15702)..(15839)
NAME/KEY: misc.feature
LOCATION: (14871)

OTHER INFORMATION: n = a or c or g or t
IS-09-608-2854-59

```

	Query Match	5.98;	Score 228.8;	DB 4;	Length 15977;	
	Best Local Similarity	75.48;	Pred.No. 1.7e-43;			
	Matches 261;	Conservative	34;	Mismatches 47;	Indels 4;	Gaps 2;
OY	1462	tgttcacaccatcttctcacccttttttttgagatgagctcgctc-tcttgcccaag	1520			
Db	12003					
OY	1521	cctggagtgcgaattggcacgctltccagctcactgcacaccccgctcttcttggttaagaacat	1580			
Db	12063	:    :    :    :    :    :    :    :    :    :    :    :				
OY	1581	tctgctgcatacagctcccgcagatcacttgggatataaggacatgtgcacacaagcccgcta	1640			
Db	12123	tctctgcctcaagctccccrgrtagcttgggattaacagcrgygcgcacacacgcycygcta	12182			
OY	1641	a-----tttgttatcttttgtagagacggggttttgcaatgcttggccagcgtgctccgac	1697			
Db	12183					
OY	1698	tcttgcattgaagatgataccatctgcctcttggcctccacacagctgtctgggattaaagccgtga	1757			
Db	12243	:    :    :    :    :    :    :    :    :    :    :    :				
OY	1758	gccacacatgccgcgaccttttctcactcttccacttccacactgtgctctcat	1803			
Db	12303	:    :    :    :    :    :    :    :    :    :    :				
		gccacacacckcyrgccttttttghtkgtkcttttttttttcht	12348			

US-08-756-506-1  
Sequence 1, Application US/08756506  
Patent No. 5905185  
GENERAL INFORMATION:  
APPLICANT: Garner, Ian  
APPLICANT: Cottingham, Ian R.  
APPLICANT: Temperley, Simon M.  
APPLICANT: Foster, Donald C.  
APPLICANT: Sprecher, Cindy A.  
APPLICANT: Prunkard, Donna E.  
TITLE OF INVENTION: PROTEIN C PRODUCTION IN TRANSGENIC  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Zymogenetics, Inc.  
STREET: 1201 Eastlake Avenue East  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98102  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/756,506  
FILING DATE:  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Sawislak, Deborah A  
REGISTRATION NUMBER: 37,438  
REFERENCE/DOCKET NUMBER: 95-28  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-442-6672  
TELEFAX: 206-442-6678  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 11725 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: join(3520..3630, 5093..5117, 5210..5347, 5450  
LOCATION: ..5584, 8253..8395, 9269..9386, 10516..11102)  
US-08-756-506-1

Query Match 5.9%; Score 227; DB 2; Length 11725;  
Best Local Similarity 85.0%; Pred. No. 3.8e-43;  
Matches 266; Conservative 0; Mismatches 45; Indels 2; Gaps 1;

QY 1469 accatatttcacatttttttttgagatgagctcgtc--ctgcccagctggag 1526  
DB 6361 ACITTCCTTTTCTTTTCTTTTGTGAGATGAGTTTCACTCTGTGTGCCAGGCTGAG 6420

QY 1527 tgcataggcaagcttcagcactgcaactcgccctctgggttcaagaattcgtc 1586  
DB 6421 TGCATATACGTGATCTCAGCTCACACACCTCCGCTCTGGATTCAGAGATTCTCCT 6480

QY 1587 gcatacagctccagctgagctgagatatacagagcatgtgcacacccagccgctaatlttg 1646  
DB 6481 GCCCAGAGCTCCGAGTAGCTGGGATTAACAGCATGGCCACACGCCCACTAATTG 6540

QY 1647 tatttttaagtagaaggggttttgcacatgtggcagagctggtctcgaactctgact 1706  
DB 6541 TGTTTTATAGTAGAAGAGGGGTTTCTCCGTGTGTGCTCAAGCTGTGAACTCTCGACT 6600

QY 1707 cagatgatccatctgcttgccctccacagtgctggatatacaggggtgagacacatg 1766  
DB 6601 CAGGTGATCCACCTGCTTGCGCTCTTAAGTGTGGGATTACAGGGGTGACCAACCGCG 6660

QY 1767 ccaggcctcttc 1779  
DB 6661 CCACAGCTCTTTC 6673

RESULT 6  
US-08-884-324-9/c  
Sequence 9, Application US/08884324  
Patent No. 6060283  
GENERAL INFORMATION:  
APPLICANT: Takamori OKURA  
APPLICANT: Kakuji TORIGOE  
APPLICANT: Masashi KURIMOTO  
TITLE OF INVENTION: GENOMIC DNA ENCODING A POLYPEPTIDE CAPABLE  
TITLE OF INVENTION: OF INDUCING THE PRODUCTION OF INTERFERON-  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROWDY AND NEIMARK  
STREET: 419 Seventh Street, N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/884,324  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 185,305/96  
FILING DATE: 27-JUN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: BROWDY, Roger L.  
REGISTRATION NUMBER: 25,618  
REFERENCE/DOCKET NUMBER: OKURA-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4773 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: Genomic DNA  
ORIGINAL SOURCE:  
ORGANISM: human  
TISSUE TYPE: placenta  
FEATURE:  
NAME/KEY: Intron  
LOCATION: 1..4773  
IDENTIFICATION METHOD: E  
US-08-884-324-9

Query Match 5.8%; Score 223.6; DB 3; Length 4773;  
Best Local Similarity 85.3%; Pred. No. 1.5e-42;  
Matches 261; Conservative 0; Mismatches 44; Indels 1; Gaps 1;

QY 1476 ttltcaacttttttgagatgagctcgtc-tcttgccaggtgagtgcaatg 1554  
DB 1673 TTTTCTTTTCTTTTGTGAGCCGAGTTTGTGTGCCCCAGGCTGAGTCAATGG 1614

QY 1535 cagcttcaagctcaacgcaactccgcctcttggtgtcaagcaatctcgtgcatcagc 1594  
DB 1613 CAGCATCTAGGCTCATGCGAACCTCTGCTCCGCCGTTCAAAAGATTCTCTGCTCAGC 1554

QY 1595 ctcccgagtaactgagattacagagcatgtgccacacgcccgciaatttgtatttta 1654

Db - 1553 CTCGCCAGTAGCTGGATTACAGCATATGCCACACGCCACGATATTGTTATTTT 1494  
QY 1655 gtagagacggggttttgcacatgttgcacggcgtgtctcgaactcttgactgaatgat 1714  
Db 1493 gtagagacggggttttgcacatgttgcacggcgtgtctcgaactcttgactgaatgat 1434  
QY 1715 ccacgtccttgccctcccaagtgctgagattacaggcgtgagccacatgccggcct 1774  
Db 1433 cccccccttcggcctcccaagtgctgagattacaggcgtgagccacatgccggcct 1374  
QY 1775 cttctc 1780  
Db 1373 CATTCT 1368

RESULT 7  
US-08-884-324-13/c  
; Sequence 13, Application US/08884324  
; Patent No. 6060283  
; GENERAL INFORMATION:  
; APPLICANT: Takamori OKURA  
; APPLICANT: Kakui TORIGOE  
; APPLICANT: Masahi KURIMOTO  
; TITLE OF INVENTION: GENOMIC DNA ENCODING A POLYPEPTIDE CAPABLE  
; TITLE OF INVENTION: OF INDUCING THE PRODUCTION OF INTERFERON-  
; NUMBER OF SEQUENCES: 35  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROWDY AND NEIMARK  
; STREET: 419 Seventh Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/884,324  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 185,305/96  
; FILING DATE: 27-JUN-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BROWDY, Roger L.  
; REGISTRATION NUMBER: 25,618  
; REFERENCE/DOCKET NUMBER: OKURA-1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-628-5197  
; TELEFAX: 202-737-3528  
; INFORMATION FOR SEQ ID NO: 13:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 11464 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: Genomic DNA  
; ORIGINAL SOURCE:  
; ORGANISM: human  
; TISSUE TYPE: placenta  
; FEATURE:  
; NAME/KEY: 5'UTR  
; LOCATION: 1..3  
; IDENTIFICATION METHOD: E  
; NAME/KEY: leader peptide  
; LOCATION: 4..82  
; IDENTIFICATION METHOD: S  
; NAME/KEY: Intron  
; LOCATION: 83..1453  
; IDENTIFICATION METHOD: E

NAME/KEY: leader peptide  
LOCATION: 1454..1465  
IDENTIFICATION METHOD: S  
NAME/KEY: Intron  
LOCATION: 1466..4848  
IDENTIFICATION METHOD: E  
NAME/KEY: leader peptide  
LOCATION: 4849..4865  
IDENTIFICATION METHOD: S  
NAME/KEY: mat peptide  
LOCATION: 4866..4983  
IDENTIFICATION METHOD: S  
NAME/KEY: Intron  
LOCATION: 4984..6317  
IDENTIFICATION METHOD: E  
NAME/KEY: mat peptide  
LOCATION: 6318..6451  
IDENTIFICATION METHOD: S  
NAME/KEY: Intron  
LOCATION: 6452..11224  
IDENTIFICATION METHOD: E  
NAME/KEY: mat peptide  
LOCATION: 11225..11443  
IDENTIFICATION METHOD: S  
NAME/KEY: 3'UTR  
LOCATION: 11444..11464  
IDENTIFICATION METHOD: E

US-08-884-324-13

Query Match 5.8%; Score 223.6; DB 3; Length 11464;  
Best Local Similarity 85.3%; Pred. No. 2,3e-42;  
Matches 261; Conservative 0; Mismatches 44; Indels 1; Gaps 1;

QY 1476 ttctcactttttttgagatgagtcctc-tctgcccagctgagtgcaatg 1534  
Db 8124 TTTTttttttttttttgagacccagtttgcctgttgcctccagctgagtgcaatg 8065  
QY 1535 caagttcctcagctcactgaacctccgctcttggttaagaactctgctcagc 1594  
Db 8064 CACGATCTAGGCTCAGTCAGCACTCTGCTCCGGGTTCAAAAGATTCTCTCCACG 8005

QY 1595 ctcccgactctgagattacagcatgtgcccacagcccgactaatctgatttta 1654  
Db 8004 CTCGCCAGTAGCTGGGATTACAGCATATGCCACACGCCACGATATTGTTATTTT 7945

QY 1655 gtagagacggggttttgcacatgttgcacggcgtgtctcgaactcttgactgaatgat 1714  
Db 7944 gtagagacggggttttgcacatgttgcacggcgtgtctcgaactcttgactgaatgat 7885

QY 1715 ccacgtccttgccctcccaagtgctgagattacaggcgtgagccacatgccggcct 1774  
Db 7884 cccccccttcggcctcccaagtgctgagattacaggcgtgagccacatgccggcct 7825

QY 1775 cttctc 1780  
Db 7824 CATTCT 7819

RESULT 8  
US-08-884-324-14/c  
; Sequence 14, Application US/08884324  
; Patent No. 6060283  
; GENERAL INFORMATION:  
; APPLICANT: Takamori OKURA  
; APPLICANT: Kakui TORIGOE  
; APPLICANT: Masahi KURIMOTO  
; TITLE OF INVENTION: GENOMIC DNA ENCODING A POLYPEPTIDE CAPABLE  
; TITLE OF INVENTION: OF INDUCING THE PRODUCTION OF INTERFERON-  
; NUMBER OF SEQUENCES: 35  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROWDY AND NEIMARK  
; STREET: 419 Seventh Street, N.W., Suite 300



```

1      CITY: Washington
2      COUNTRY: D.C.
3      ZIP: 20004
4
5      COMPUTER READABLE FORM:
6      MEDIUM TYPE: floppy disk
7      COMPUTER: IBM PC compatible
8      OPERATING SYSTEM: PC-DOS/MS-DOS
9      SOFTWARE: Patent In Release #1.0, Version #1.30
10     CURRENT APPLICATION DATA:
11     APPLICATION NUMBER: US/08/884,324
12     FILING DATE:
13     CLASSIFICATION: 435
14     PRIOR APPLICATION DATA:
15     APPLICATION NUMBER: JP 185,305/96
16     FILING DATE: 27-JUN-1996
17     ATTORNEY/AGENT INFORMATION:
18     NAME: BROWDY, Roger L.
19     REGISTRATION NUMBER: 25,618
20     REFERENCE/DOCKET NUMBER: OKURA-1
21     TELECOMMUNICATION INFORMATION:
22     TELEPHONE: 202-628-5197
23     TELEFAX: 202-737-3528
24     INFORMATION FOR SEQ ID NO: 14:
25     SEQUENCE CHARACTERISTICS:
26     LENGTH: 28994 base pairs
27     TYPE: nucleic acid
28     STRANDEDNESS: double
29     TOPOLOGY: linear
30     MOLECULE TYPE: Genomic DNA
31     ORIGINAL SOURCE:
32     ORGANISM: human
33     TISSUE TYPE: Placenta
34
35     FEATURE:
36     NAME/KEY: 5'UTR
37     LOCATION: 1..15606
38     IDENTIFICATION METHOD: E
39     NAME/KEY: leader peptide
40     LOCATION: 15607..15685
41     IDENTIFICATION METHOD: S
42     NAME/KEY: Intron
43     LOCATION: 15686..17056
44     IDENTIFICATION METHOD: E
45     NAME/KEY: leader peptide
46     LOCATION: 17057..17068
47     IDENTIFICATION METHOD: S
48     NAME/KEY: Intron
49     LOCATION: 17069..20451
50     IDENTIFICATION METHOD: E
51     NAME/KEY: leader peptide
52     LOCATION: 20452..20468
53     IDENTIFICATION METHOD: S
54     NAME/KEY: mat peptide
55     LOCATION: 20469..20586
56     IDENTIFICATION METHOD: S
57     NAME/KEY: Intron
58     LOCATION: 20587..21920
59     IDENTIFICATION METHOD: E
60     NAME/KEY: mat peptide
61     LOCATION: 21921..22054
62     IDENTIFICATION METHOD: S
63     NAME/KEY: Intron
64     LOCATION: 22055..26827
65     IDENTIFICATION METHOD: E
66     NAME/KEY: mat peptide
67     LOCATION: 26828..27046
68     IDENTIFICATION METHOD: S
69     NAME/KEY: 3'UTR
70     LOCATION: 27047..28994
71     IDENTIFICATION METHOD: E
72
73     IS-08-884-324-14

```



Query Match 5.78; Score 221.8; DB 4; Length 15297

Patent No. 05945522  
 GENERAL INFORMATION:  
 APPLICANT: Cohen, Daniel  
 APPLICANT: Chumakov, Ilya  
 APPLICANT: Blumenfeld, Marta  
 APPLICANT: Bougueleret, Lydie  
 TITLE OF INVENTION: Prostate cancer gene  
 NUMBER OF SEQUENCES: 68  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Knobb, Martens, Olson & Bear  
 STREET: 501 West Broadway  
 CITY: San Diego  
 STATE: California  
 COUNTRY: USA  
 ZIP: 92101-3505  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: Win95  
 SOFTWARE: Word  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/996,306  
 FILING DATE:  
 CLASSIFICATION:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Israelsen, Ned A.  
 REGISTRATION NUMBER: 29,655  
 REFERENCE/DOCKET NUMBER: GENSET.018A  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (619) 235-8550  
 TELEFAX: (619) 235-0176  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 56516 base pairs  
 TYPE: NUCLEIC ACID  
 STRANDEDNESS: DOUBLE  
 TOPOLOGY: LINEAR  
 MOLECULE TYPE: GENOMIC DNA  
 ORIGINAL SOURCE:  
 ORGANISM: Homo sapiens  
 FEATURE:  
 NAME/KEY: Promoter

1	LOCATION:	1629..1870	PROSCAN
2	IDENTIFICATION METHOD:		
3	FEATURE:		
4	NAME/KEY:	Potential ATG	
5	LOCATION:	1998..2000	
6	FEATURE:		
7	NAME/KEY:	Exon 1	
8	LOCATION:	2001..2216	
9	FEATURE:		
10	NAME/KEY:	ATG	
11	LOCATION:	2031..2033	
12	FEATURE:		
13	NAME/KEY:	Tyr phos	
14	LOCATION:	11694..14332	
15	FEATURE:		
16	NAME/KEY:	SEQ ID42	
17	LOCATION:	11930..11947	
18	FEATURE:		
19	NAME/KEY:	SEQ ID24	
20	LOCATION:	12057..12103	
21	FEATURE:		
22	NAME/KEY:	SEQ ID51	
23	LOCATION:	comp1(12339..12358)	
24	FEATURE:		
25	NAME/KEY:	SEQ ID64	
26	LOCATION:	13547..13564	
27	FEATURE:		
28	NAME/KEY:	SEQ ID58	
29	LOCATION:	13657..13703	
30	FEATURE:		
31	NAME/KEY:	SEQ ID67	
32	LOCATION:	comp1(13962..13981)	
33	FEATURE:		
34	NAME/KEY:	Exon 2	
35	LOCATION:	18196..18265	
36	FEATURE:		
37	NAME/KEY:	Exon 3	
38	LOCATION:	23717..23832	
39	FEATURE:		
40	NAME/KEY:	Exon 4	
41	LOCATION:	25571..25660	
42	FEATURE:		
43	NAME/KEY:	SEQ ID43	
44	LOCATION:	34216..34234	
45	FEATURE:		
46	NAME/KEY:	SEQ ID25	
47	LOCATION:	34469..34515	
48	FEATURE:		
49	NAME/KEY:	SEQ ID52	
50	LOCATION:	comp1(34625..34645)	
51	FEATURE:		
52	NAME/KEY:	Exon 5	
53	LOCATION:	34669..34759	
54	FEATURE:		
55	NAME/KEY:	Exon 6	
56	LOCATION:	40688..40846	
57	FEATURE:		
58	NAME/KEY:	Exon 7	
59	LOCATION:	48070..48193	
60	FEATURE:		
61	NAME/KEY:	Exon 8	
62	LOCATION:	50182..54523	
63	FEATURE:		
64	NAME/KEY:	SEQ ID65	
65	LOCATION:	51149..51168	
66	FEATURE:		
67	NAME/KEY:	SEQ ID59	
68	LOCATION:	51448..51494	
69	FEATURE:		
70	NAME/KEY:	SEQ ID68	
71	LOCATION:	comp1(51482..51499)	
72	FEATURE:		
73	NAME/KEY:	SEQ ID44	

```

: LOCATION: 51596..51613
:
: FEATURE:
:
: NAME/KEY: SEQ ID26
: LOCATION: 51612..51658
:
: FEATURE:
:
: NAME/KEY: SEQ ID53
: LOCATION: comp151596..52015
:
: FEATURE:
:
: NAME/KEY: polyad signal
: LOCATION: 54445..54450
:
: OS-08-996-306-1

```

Query Match	5.7%;	Score 220.2;	DB 2;	Length 56516;
Best Local Similarity	76.5%;	Pred. No. 3.3e-41;		
Matches 270; Conservative	0;	Mismatches 83;	Indels 0;	Gaps 0;

QY	1435	tttcctactctcaacatcgcacacccaatgataccatctcttccacatttttgg	1494
Db	20869	ttttcttcccccttaacctttaaagagatctacttttgacaacacttttgcccttttttttttga	20728
QY	1495	gattgagatctgcgtctctcttgcacaggtcgtgagtcaatgacagtcttcagctcaetgca	1554
Db	20729	gattgagatttgcactcttctgtttgtgccacagcctgggtgttaaggatgacaaattccagctcacatgcga	20788
QY	1555	acctccgcctcttggtttcaagcaattctgctgcgaacagctcccgagatccctggagta	1614
Db	20789	accttttgccctcccgsgtttaaacctatttttccgtccctcagcctcccgagatgctggagattta	20844
QY	1615	caagcatgtgccaacacagcccggtcaatttggatattttagtagagacggggttttgcca	1674
Db	20849	cagacatgcaacacacacgcctctggccttaattttgtatttttgcattgacatgagatggggtttcaacca	20908
QY	1675	tgattgacagagctggtctcgaaccttgaactcagatgacatgcactgtccttggccctccca	1734
Db	20909	tgttttgatcagagctgggtcttgaaactcctgacctcagatgatctgcctctgcctcctggcctcccca	20966
QY	1735	caagtgtcgtgattacaagcgtgtgagccaacacagcccgagctcttctcaacctt	1787
Db	20969	aactgtcagagattttacagcgttatgagccactgtgcccctctgctctctatgtttttggcctttt	21021

```

RESULT 14
US-09-338-907-1
Sequence 1, Application US/09338907
Patent No. 6265546
GENERAL INFORMATION:
APPLICANT: Cohen, Daniel
APPLICANT: Blumenfeld, Marta
APPLICANT: Ilya, Chumakov
APPLICANT: Bougueteleret, Lydie
TITLE OF INVENTION: PROSTATE CANCER GENE
FILE REFERENCE: GENSET.18C1CIP
CURRENT APPLICATION NUMBER: US/09/338,907
EARLIER FILING DATE: 1999-06-23
EARLIER APPLICATION NUMBER: 08/996,306
EARLIER FILING DATE: 1997-12-22
EARLIER APPLICATION NUMBER: 60/099,658
EARLIER FILING DATE: 1998-09-09
EARLIER APPLICATION NUMBER: 09/218,207
EARLIER FILING DATE: 1998-12-22
NUMBER OF SEQ. ID NOS: 578
SOFTWARE: Patent.pm
SEQ. ID NO 1
LENGTH: 56516
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: promoter
LOCATION: 1629..1870
OTHER INFORMATION: Identification method
NAME/KEY: misc.feature
Proscan

```

```

LOCATION: 1998..2000
OTHER INFORMATION: potential start codon
FEATURE:
NAME/KEY: exon
LOCATION: 2001..2216
OTHER INFORMATION: exon1
FEATURE:
NAME/KEY: misc-feature
LOCATION: 2031..2033
OTHER INFORMATION: ATG
FEATURE:
NAME/KEY: misc-feature
LOCATION: 11694..14332
OTHER INFORMATION: Tyr Phos
FEATURE:
NAME/KEY: primer_bind
LOCATION: 11930..11947
OTHER INFORMATION: upstream amplification primer 4-77 SEQ ID42
FEATURE:
NAME/KEY: allele
LOCATION: 12057..12103
OTHER INFORMATION: polymorphic fragment 4-77 SEQ ID24
FEATURE:
NAME/KEY: primer_bind
LOCATION: 12339..12358
OTHER INFORMATION: downstream amplification primer 4-77 SEQ ID51, complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 13547..13564
OTHER INFORMATION: upstream amplification primer 4-73 SEQ ID64
FEATURE:
NAME/KEY: allele
LOCATION: 13657..13703
OTHER INFORMATION: polymorphic fragment 4-73 SEQ ID58
FEATURE:
NAME/KEY: primer_bind
LOCATION: 13962..13981
OTHER INFORMATION: downstream amplification primer 4-73 SEQ ID67, complement
FEATURE:
NAME/KEY: exon
LOCATION: 18196..18265
OTHER INFORMATION: exon 2
FEATURE:
NAME/KEY: exon
LOCATION: 23717..23832
OTHER INFORMATION: exon 3
FEATURE:
NAME/KEY: exon
LOCATION: 25571..25660
OTHER INFORMATION: exon 4
FEATURE:
NAME/KEY: primer_bind
LOCATION: 34216..34234
OTHER INFORMATION: upstream amplification primer 99-217 SEQ ID43
FEATURE:
NAME/KEY: allele
LOCATION: 34469..34515
OTHER INFORMATION: polymorphic fragment 99-217 SEQ ID25
FEATURE:
NAME/KEY: primer_bind
LOCATION: 34625..34645
OTHER INFORMATION: downstream amplification primer 99-217 SEQ ID52, complement
FEATURE:
NAME/KEY: exon
LOCATION: 34669..34759
OTHER INFORMATION: exon 5
FEATURE:
NAME/KEY: exon
LOCATION: 40688..40846
OTHER INFORMATION: exon 6
FEATURE:
NAME/KEY: exon
LOCATION: 48070..48193

```

```

OTHER INFORMATION: exon 7
FEATURE:
NAME/KEY: exon
LOCATION: 50182..54523
OTHER INFORMATION: exon 8
FEATURE:
NAME/KEY: primer_bind
LOCATION: 51149..51168
OTHER INFORMATION: upstream amplification primer 4-65 SEQ ID65
FEATURE:
NAME/KEY: allele
LOCATION: 51448..51494
OTHER INFORMATION: polymorphic fragment 4-65 SEQ ID59
FEATURE:
NAME/KEY: primer_bind
LOCATION: 51482..51499
OTHER INFORMATION: downstream amplification primer 4-65 SEQ ID68, complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 51596..51613
OTHER INFORMATION: upstream amplification primer 4-67 SEQ ID44
FEATURE:
NAME/KEY: allele
LOCATION: 51612..51658
OTHER INFORMATION: polymorphic fragment 4-67 SEQ ID26
FEATURE:
NAME/KEY: primer_bind
LOCATION: 51996..52015
OTHER INFORMATION: downstream amplification primer 4-67 SEQ ID53, complement
FEATURE:
NAME/KEY: polyA-signal
LOCATION: 54445..54450
OTHER INFORMATION: AATAAA
US-09-338-907-1

```

```

Query Match          5.7%; Score 220.2; DB 4; Length 56516;
Best Local Similarity 76.5%; Pred. No. 3.3e-41;
Matches 270; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 1435 ttccatcttcacacgtgcacccatcatgttccacactatcttcacacttttttga 1494
D 20669 ttctcccttatccctcaacgaatccactttgacacacacttcttttttga 20728
QY 1495 gatggagctcgcctctcttgcacagcttgagtgcaatggcagcttcacactatgca 1554
D 20729 gatggagcttcacactctgttgcacagcttggtgtaatgtgcaatcgcactacgca 20788
QY 1555 acctcgccctctgggttcaagcaattctgctgcatcagccctccgagctcggagta 1614
D 20789 accttgcctccggttcaagcaatttctcctcagcctccgagctgagta 20848
QY 1615 caggcatgtgcacacacagcccggttaatttgatttgtagagcggtttgca 1674
D 20849 cagacatgcacacacacagcccggttaatttgatttgtagagcggtttgca 20908
QY 1675 ttttggcagcgctgtctgcgaactcttgaactcaagatgacatcctctggcctcca 1734
D 20909 ttttggcagcgctgtctgttaactctgcgaactcaagatgactcctggcctcccca 20968
QY 1735 cagtgctggattacagcggtgacacacatgcccggtccttctcactt 1787
D 20969 aagtgaggtattacagcggtgacacacatgcccggtccttctcactt 21021

```

```

RESULT 15
US-09-218-207-1
; Sequence 1, Application US/09218207
; Patent No. 6346381
; GENERAL INFORMATION:
; APPLICANT: Cohen, Daniel
; APPLICANT: Blumenfeld, Marica
; APPLICANT: Ilya, Chumakov

```

```
APPLICANT: Bougueleret, Lydie
TITLE OF INVENTION: Prostate cancer gene
FILE REFERENCE: GENSET.018Cp1
CURRENT APPLICATION NUMBER: US/09/218,207
CURRENT FILING DATE: 1998-12-22
EARLIER APPLICATION NUMBER: 08/996,306
EARLIER FILING DATE: 1997-12-22
EARLIER APPLICATION NUMBER: 60/099,658
EARLIER FILING DATE: 1998-09-09
NUMBER OF SEQ ID NOS: 578
SOFTWARE: Patent.pm
SEQ ID NO 1
LENGTH: 56516
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: promoter
LOCATION: 1629..1870
OTHER INFORMATION: Identification method Proscan
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1998..2000
OTHER INFORMATION: potential start codon
FEATURE:
NAME/KEY: exon
LOCATION: 2001..2216
OTHER INFORMATION: exon1
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2031..2033
OTHER INFORMATION: ATG
FEATURE:
NAME/KEY: misc_feature
LOCATION: 11694..14332
OTHER INFORMATION: Tyr Phos
FEATURE:
NAME/KEY: primer_bind
LOCATION: 11930..11947
OTHER INFORMATION: upstream amplification primer 4-77 SEQ ID42
FEATURE:
NAME/KEY: allele
LOCATION: 12057..12103
OTHER INFORMATION: polymorphic fragment 4-77 SEQ ID24
FEATURE:
NAME/KEY: primer_bind
LOCATION: 12339..12358
OTHER INFORMATION: downstream amplification primer 4-77 SEQ ID51, complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 13547..13564
OTHER INFORMATION: upstream amplification primer 4-73 SEQ ID64
FEATURE:
NAME/KEY: allele
LOCATION: 13657..13703
OTHER INFORMATION: polymorphic fragment 4-73 SEQ ID58
FEATURE:
NAME/KEY: primer_bind
LOCATION: 13962..13981
OTHER INFORMATION: downstream amplification primer 4-73 SEQ ID67, complement
FEATURE:
NAME/KEY: exon
LOCATION: 18196..18265
OTHER INFORMATION: exon 2
FEATURE:
NAME/KEY: exon
LOCATION: 23717..23832
OTHER INFORMATION: exon 3
FEATURE:
NAME/KEY: exon
LOCATION: 25571..25660
OTHER INFORMATION: exon 4
FEATURE:
NAME/KEY: primer_bind
```

```
LOCATION: 34216..34234
OTHER INFORMATION: upstream amplification primer 99-217 SEQ ID43
FEATURE:
NAME/KEY: allele
LOCATION: 34469..34515
OTHER INFORMATION: polymorphic fragment 99-217 SEQ ID25
FEATURE:
NAME/KEY: primer_bind
LOCATION: 34625..34645
OTHER INFORMATION: downstream amplification primer 99-217 SEQ ID52, complement
FEATURE:
NAME/KEY: exon
LOCATION: 34669..34759
OTHER INFORMATION: exon 5
FEATURE:
NAME/KEY: exon
LOCATION: 40688..40846
OTHER INFORMATION: exon 6
FEATURE:
NAME/KEY: exon
LOCATION: 48070..48193
OTHER INFORMATION: exon 7
FEATURE:
NAME/KEY: exon
LOCATION: 50182..54523
OTHER INFORMATION: exon 8
FEATURE:
NAME/KEY: primer_bind
LOCATION: 51149..51168
OTHER INFORMATION: upstream amplification primer 4-65 SEQ ID65
FEATURE:
NAME/KEY: allele
LOCATION: 51448..51494
OTHER INFORMATION: polymorphic fragment 4-65 SEQ ID59
FEATURE:
NAME/KEY: primer_bind
LOCATION: 51482..51499
OTHER INFORMATION: downstream amplification primer 4-65 SEQ ID65
FEATURE:
NAME/KEY: primer_bind
LOCATION: 51596..51613
OTHER INFORMATION: upstream amplification primer 4-67 SEQ ID44
FEATURE:
NAME/KEY: allele
LOCATION: 51612..51658
OTHER INFORMATION: polymorphic fragment 4-67 SEQ ID26
FEATURE:
NAME/KEY: primer_bind
LOCATION: 51996..52015
OTHER INFORMATION: downstream amplification primer 4-67 SEQ ID53, complement
FEATURE:
NAME/KEY: polyA_signal
LOCATION: 54445..54450
OTHER INFORMATION: AATTAAA
US-09-218-207-1
```

```
Query Match 5.7%; Score 220.2; DB 4; Length 56516;
Best Local Similarity 76.5%; Pred. No. 3.3e-41;
Matches 270; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 1435 ttccatcttcacacgtccacactatgttcaacctactcttcacattttttga 1494
DB 20669 ttcttcctccctatctccttaacagatcttctgaacaacatttctttttga 20728
QY 1495 gatgagatctgcctctctgcccaggctgagtgacaatgccaagtctcagctcactga 1554
DB 20729 gatgagatctcactctctgttgcctcagctggtgtaagtgtgaatcctcagctcactga 20788
QY 1555 acctccgctcttgggttcaagaactctgctgcacacgtcccgagtaacctgggatta 1614
DB 20789 accttgcctcccggttcaagcaatttctcgtccctcagcctcccgagtaagtggatta 20848
```

Sat Jun 8 10:13:30 2002

us-09-729-674-1.rni

Page 12

[illegible]

```
Search completed: June  8, 2002, 00:10:10
Job time: 11430 sec
```



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 8, 2002, 02:08:55 ; Search time 89.01 Seconds  
(without alignments)  
779.361 Million cell updates/sec

Title: US-09-729-674-2

Perfect score: 2130

Sequence: 1 MEGRGGAALALLLCCVACA.....DDEGEVEYVDLLELEETS 401

Scoring table:

BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 10%

Listing first 45 summaries

Database :

SPTREMBL\_19:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mmc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_protist:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacterioph:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1038	48.7	222	11 Q9D8W2	Q9D8W2 mus musculus
2	572.5	26.9	804	4 Q96SN3	Q96SN3 homo sapien
3	571	26.8	736	4 Q96SL5	Q96SL5 homo sapien
4	571	26.8	736	4 Q96SK8	Q96SK8 homo sapien
5	567	26.6	736	4 Q9HC86	Q9HC86 homo sapien
6	545	25.6	742	11 Q9CWS5	Q9CWS5 mus musculus
7	539	25.3	363	4 Q9HC87	Q9HC87 homo sapien
8	534.5	25.1	728	11 Q9R1J8	Q9R1J8 ratius norv
9	534.5	25.0	542	11 Q9QZT6	Q9QZT6 mus musculus
10	532.5	25.0	747	11 Q9QZT7	Q9QZT7 mus musculus
11	372	17.5	476	5 Q9NA33	Q9NA33 caenorhabdi
12	258.5	12.1	527	4 Q9NV12	Q9NV12 homo sapien
13	161.5	7.6	545	11 Q98836	Q98836 mus musculus
14	161.5	7.6	551	4 Q15740	Q15740 homo sapien
15	159.5	7.5	551	4 Q13512	Q13512 homo sapien
16	118.5	5.6	598	5 Q9BPP2	Q9BPP2 caenorhabdi

17	104	4.9	317	16 Q97R54	Q97R54 streptococc
18	104	4.9	853	16 Q67796	Q67796 aquifex aeo
19	100.5	4.7	688	12 Q91TS4	Q91TS4 lupala hep
20	100	4.7	1006	5 Q97090	Q97090 drosophila
21	100	4.7	1006	5 Q9VUC7	Q9VUC7 drosophila
22	99	4.6	1388	6 Q9GL21	Q9GL21 canis fami
23	98.5	4.6	1575	16 Q9P0H4	Q9P0H4 ureaplasma
24	98	4.6	455	5 Q16626	Q16626 caenorhabdi
25	98	4.6	484	16 Q06968	Q06968 bacillus su
26	98	4.6	4202	11 Q91V63	Q91V63 mus musculu
27	98	4.6	4731	11 Q91XP8	Q91XP8 mus musculu
28	97.5	4.6	910	3 Q59796	Q59796 schizosacch
29	97.5	4.6	1162	16 Q9PMG0	Q9PMG0 campylobact
30	96	4.5	788	17 Q58603	Q58603 pyrococcus
31	96	4.5	3477	11 Q9NMH8	Q9NMH8 mus musculu
32	96	4.5	4202	11 Q91XP9	Q91XP9 mus musculu
33	96	4.5	4731	11 Q91X00	Q91X00 mus musculu
34	95.5	4.5	240	17 Q981E0	Q981E0 sulfolobus
35	95.5	4.5	336	5 Q95OS2	Q95OS2 caenorhabdi
36	94	4.4	435	16 Q9PMV0	Q9PMV0 campylobact
37	94	4.4	439	5 Q9GSD0	Q9GSD0 plasmidium
38	93.5	4.4	541	5 Q95P11	Q95P11 brugia mala
39	93	4.4	434	16 Q97KG2	Q97KG2 clostridium
40	93	4.4	757	10 Q9F174	Q9F174 arabidopsis
41	92.5	4.3	729	16 Q9PHP5	Q9PHP5 campylobact
42	92	4.3	411	16 Q9HV83	Q9HV83 pseudomonas
43	92	4.3	534	3 Q9USM8	Q9USM8 schizosacch
44	92	4.3	877	2 Q9RG09	Q9RG09 neisseria m
45	91.5	4.3	429	2 Q91AL2	Q91AL2 moraxella c

## ALIGNMENTS

RESULT	ID	PRELIMINARY	PRT	222 AA.
Q9D8W2	Q9D8W2	Q9D8W2		
AC	01-JUN-2001 (TREMBLrel. 17, Created)			
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE	5730529N23RIK PROTEIN.			
OS	CRTPP OR 5730529N23RIK.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=PANCREAS;			
RX	MEDLINE=21085660; PubMed=11217851;			
RA	Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,			
RA	Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,			
RA	Alzawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamada I.,			
RA	Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,			
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,			
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochias H.,			
RA	Kuethl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,			
RA	Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,			
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,			
RA	Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,			
RA	Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,			
RA	Gustincich S., Hill D., Hofmann M., Hume D.A., Kamya M., Lee N.H.,			
RA	Lyons P., Marchionni L., Mashima J., Mazzarelli J., Kombaris P.,			
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,			
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,			
RA	Suzuki H., Toyokawa K., Wang K.H., Weitz C., Whitaker C., Wilming L.,			
RA	Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,			
RA	Hayashizaki Y.,			
RT	"Functional annotation of a full-length mouse cDNA collection."			
RL	Nature 409:685-690(2001).			
DR	EMBL, AK007624, BAB25144.1; -			
DR	MGD, MGI:1891221, Crtpap.			

```

SQ      SEQUENCE      222 AA;  25803 MW;  336DF2019A126B04 CRC64;

```

Query Match	48.7%	Score 1038	DB 11	Length 222
Best Local	86.5%	Pred. No. 2.4e81		
Matches 192	Conservative 13	Mismatches 17	Indels 0	Gaps 0

[illegible]

RESULT	2			
ID	096SN3	PRELIMINARY:	PRT:	804 AA.
AC	096SN3			
DT	01-DEC-2001 (TrEMBLrel. 19, Created)			
DT	01-DEC-2001 (TrEMBLrel. 19, last sequence update)			
DT	01-DEC-2001 (TrEMBLrel. 19, last annotation update)			
DE	CNNA FL114742 FIS. CLONE NT2P3002650, HIGHLY SIMILAR TO MS			
DE	MUSCULUS GROWTH SUPPRESSOR 1L (GROSL) MRNA.			
OS	Homo sapiens (human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RY	[1]			
RP	SEQUENCE FROM N.A.			
RA	Isoegai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,			
RA	Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,			
RA	Nagatsuna M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara			
RA	Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi			
RA	Yamanoto J., Kimura K., Murakami K., Ishi S., Kawai Y., Saito K			
RA	Nishimura K., Iwawano T.,			
RT	"MEDO human cDNA sequencing project."			
RL	Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.			
RL	EMBL: AK027648; BAB55264.1; "			
SO	SEQUENCE 804 AA; 90544 MW; 23C989DE603D208CD CRC64;			

Query Match	26.98;	Score 572.5;	DB 4;	Length 804;
Best Local Similarity	35.38;	Pred. No. 1.3e-40;		
Matches 145;	Conservative 65;	Mismatches 158;	Indels 43;	Gaps 12;

QY 6 RGAALLLALLLVACALAGRAOYERYEFSRPELBELPLESAVNHADKSGEMESXG 65  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
Db 4 RALKLTTLTLLAVAA--ASQAIVESEAGWGMYTPDLL-----FAEGTAATAARDWPCVYL 56  
  
QY 66 YLEISLRHLRLDRDSEAFCHRNCSA-----APQEPBAGLASYPELRLFGILLR 115  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
Db 57 SMERALSRSRALRLRLRCRTOCOADPWEILDPMDSPPAQASMAAILRDLSPFEGILLR 116  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
  
QY 116 AHCLCKROGQIPARQSQPSREVLADDFORREPKYFLQFAFKNNNLPKATIAAHTEFLKH 175  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
Db 117 AACLRRC-LGPRAHSL--SEEMLETERKSSPVNYLOVAAYFKINKLEKAAMAAHTFPVCN 173  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
  
QY 176 PDDEMRNNAYYKSLPGAE DY -TKDETLESYSLEIFIRAYANGENWRYSIDMETALP 234  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
Db 174 PEHMENQNDIYQGTMSGVKEADPKDELTFPHNQEFRLTGRLITSBDPRQCAVYHNLEALQ 233  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
  
QY 235 DFFEAFTCELAACGSSEIKDFK-----DPVLSIADHYVELECKIQCEENT-TPVIIG 287

[illegible]

RESULT	3		
096SL5			
ID	096SL5	PRELIMINARY;	PRT; 736 AA.
AC	096SL5;		
DT	01-DEC-2001 (TrEMBLrel. 19, Created)		
DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)		
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)		
DE	CDNA FLJ14774 F1S, CLONE NT2RP4000051, WEAKLY SIMILAR TO SYNAPTONEMAL COMPLEX PROTEIN SC65.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Isoegal T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.		
RA	Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,		
RA	Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu		
RA	Nakamura Y., Nagahari K., Masuno Y., Sasaki N.;		
RT	"NEDO human cDNA sequencing project."		
RL	Submitted (Mar-2001) to the EMBL/Genbank/DBJ databases.		
SR	EMBL; AK027680; BAB55291.1; ..		
SO	SEQUENCE 736 AA; 83411 MW; EA1819938FAE685E CRC64;		

Query Match	26.8%	Score 571	DB 4	length 736
Best Local Similarity	34.4%	Pred. No. 1.5e+40		
Matches 145	Conservative	68	Mismatches 168	Indels 40
				Gaps 11

QY	6	RGAAAILALILCVACALIRAGRAQOYERYSFRSPRDELMPLESAYNHALDKSGEHMAESXG	65
Db	4	RAKLKLTLLVAVVA--ASQAEEVSEAGWCVTPDDL-----FAGETAAVARGDWPCVVL	56
QY	66	YLELSLRHLRLRDSSEAFCHNCSA-----APQEPAGGLASTPELRYFGGLRR	115
Db	57	SMERALRSRAALIRLRLRLCRFCROCAADFPWELRDPWMSPPMAQSAALRLSPFGGLRR	116
QY	116	AHCLKRCOGAPAFRQSQPSREVLADFORREPYKFLQPAFKANLRLKALAAATFLKH	175
Db	117	AAALIRRC-LGPPAAHSL--SEEMELERKRKSPRYTLOYATFKIKLKAANAATFEVGN	173
QY	176	PDDMMKRNMAUYKSLPGAEY-IKDELTSYESLFTRAVAYNGENMRTSITDMLAP	234
Db	174	PEHNMQONILDYQUTMSGVKAEADKLETOHMFQRLGVRLEYSEQPOEAVPRLAALQ	233
QY	235	DEFAFYUCLAAAGCSRIKRFK-----DEYLSADHYVEULECKLOCEBNL-TPYIGG	287
Db	234	EYFAVEECRALCGSPYDYDGYNLLEYNADLFQALITHTYIOLVILCKONCVTELAHSRSRE	293
QY	288	YRVEKEFVATMYHYLQFAVYUKLNDLKNAPCAVSYLEPDDQDKVMQOMLVYUYQHRTWGL	347
Db	294	KRPDELPLSHNYLQFATVYNYNGNTQVAVEACKLLEFFRPDEVMNQMLAYTAA--L	348
QY	348	SDEHFO--PRPEAVQFPNVTTLQKELYDRAKE---NINDDEGEVVEYVDLLLEET	400
Db	349	GEENHRSIGPRESAKEVYRORSLEKEILFFAYVDVFGPIRFPDPSWTPBEVYIPKRLQEKQ	408
QY	401	S 401	
Db	409	S 409	

```

RESULT 4
O96SK8 PRELIMINARY: PRT: 736 AA.
AC 096SK8;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DE CDNA FL114791 FIS. CLONE NT2RP4001064, WEAKLY SIMILAR TO
DE SYNATONEML COMPLEX PROTEIN SC65.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA Nakamura Y., Nagahari K., Masubo Y., Sasaki N.;
RT "NEO human cDNA sequencing project."
RT Submitted (MAY-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: AK027697; BAB5305.1; -
SQ SEQUENCE 736 AA; 83393 MW; EA1909828FAE685E CRC64;

Query Match 26.8%; Score 571; DB 4; Length 736;
Best Local Similarity 34.4%; Pred. No. 1.5e-40;
Matches 145; Conservative 68; Mismatches 168; Indels 40; Gaps 11;

OY 6 RGAALLLALCVACALRAGROYERSFRSPDELMPLSAVRHNLDKYSGEHMAESXG 65
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
   4 RALKLTLLAVVAA--ASQAEVSEAGMGWTPDLL-----FAEGTAAYARGDMPGVVL 56
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 66 YLEISLRHLRLDSEAFCHRNCSA-----APQEPAGLASYPELRFGGLLR 115
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 57 SMERALRSRALRALRLRCTGCAADFPWELDPDWSPPAQASGALRLDSFGGLLR 116
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 116 ARLCKRCKGLPAPROSOPREVLADFORREPKFLQFAFKANNLPKATAAHTFLKH 175
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 117 AALRLRC-LGPPAHSL--SEEMLEFRKRSPPNYLQVAYFKINKLEKAAVAHAHTFVGN 173
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 176 PDDEMKRNNAVYKSLPGAEDY- IKDLETSYSLSFIRAVRANGENWRTSITMELALP 234
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 174 PEMEMQONLDYQTMGSGVEADFKDLETPHMOEFLGRVLSSEDPQCAVPHLEALQ 233
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 235 DFKATYECLACEGSREIKDFK-----DFYLSADHYVEVECKIQCEENL-TPYIGG 287
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 234 EYFVAEEECALCEGPDYDGYNYLEYNADLFQAITDHYIQLNCKONCVTELASHPSRE 293
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 288 YPEKEFVAITMYHLOFAYYKLLNDLKNAPCAVSYLLFDONDKYMOONLYYQYHRDTWGL 347
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 294 KPREDLPSHYNYLQFAYYNIGNYTQAVECAKTYLLFFPNDEVNOMLAYAAM-----L 348
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 348 SDEHFO--PRPAVOFENVTTLOKELYDFAKE---NIMDDGEVEVEVDLLELEET 400
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 349 GEEHTRSIGRESAKERYQRSLEKELLEFPAYDVFGIPVDPDPSWPEEVIPIRLDEKQK 408
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 401 S 401
DB 409 S 409

RESULT 5
O9HC86 PRELIMINARY: PRT: 736 AA.
AC 09HC86;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DE GROS1-L PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE-TESTIS;
RC MEDLINE=20406537; PubMed=10951563;
RA Kaul S.C., Sugihara T., Yoshida A., Nomura H., Wadhwa R.;
RT "Gros1, a potential growth suppressor on chromosome 1: its identity to
RT basement membrane-associated proteoglycan, leprecan."
RT Oncogene 19:3576-3583(2000).
DR EMBL: AF097432; AAG31019.1; -
DR InterPro: IPR000886; ER_target.
DR PROSITE: PS00014; ER_TARGET; UNKNOWN.1.
SQ SEQUENCE 736 AA; 83413 MW; 59D1205DB76ADB50 CRC64;

Query Match 26.6%; Score 567; DB 4; Length 736;
Best Local Similarity 34.2%; Pred. No. 3.4e-40;
Matches 144; Conservative 68; Mismatches 169; Indels 40; Gaps 11;

OY 6 RGAALLLALCVACALRAGROYERSFRSPDELMPLSAVRHNLDKYSGEHMAESXG 65
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
   4 RALKLTLLAVVAA--ASQAEVSEAGMGWTPDLL-----FAEGTAAYARGDMPGVVL 56
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 66 YLEISLRHLRLDSEAFCHRNCSA-----APQEPAGLASYPELRFGGLLR 115
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 57 SMERALRSRALRALRLRCTGCAADFPWELDPDWSPPAQASGALRLDSFGGLLR 116
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 116 ARLCKRCKGLPAPROSOPREVLADFORREPKFLQFAFKANNLPKATAAHTFLKH 175
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 117 AALRLRC-LGPPAHSL--SEEMLEFRKRSPPNYLQVAYFKINKLEKAAVAHAHTFVGN 173
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 176 PDDEMKRNNAVYKSLPGAEDY- IKDLETSYSLSFIRAVRANGENWRTSITMELALP 234
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 174 PEMEMQONLDYQTMGSGVEADFKDLETPHMOEFLGRVLSSEDPQCAVPHLEALQ 233
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 235 DFKATYECLACEGSREIKDFK-----DFYLSADHYVEVECKIQCEENL-TPYIGG 287
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 234 EYFVAEEECALCEGPDYDGYNYLEYNADLFQAITDHYIQLNCKONCVTELASHPSRE 293
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 288 YPEKEFVAITMYHLOFAYYKLLNDLKNAPCAVSYLLFDONDKYMOONLYYQYHRDTWGL 347
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 294 KPREDLPSHYNYLQFAYYNIGNYTQAVECAKTYLLFFPNDEVNOMLAYAAM-----L 348
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 348 SDEHFO--PRPAVOFENVTTLOKELYDFAKE---NIMDDGEVEVEVDLLELEET 400
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 349 GEEHTRSIGRESAKERYQRSLEKELLEFPAYDVFGIPVDPDPSWPEEVIPIRLDEKQK 408
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 401 S 401
DB 409 S 409

RESULT 6
O9CWK5 PRELIMINARY: PRT: 742 AA.
AC 09CWK5;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DE 2410024C15RIK PROTEIN.
GN GROS1 OR 2410024C15RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=C57BL/6J; TISSUE=EMBRYONIC STEM CELLS;
RC MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,

```

RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batilov S., Casavant T.,  
 RA Pleischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,  
 RA Knehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,  
 RA Schirml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Futuno M., Kono H., Baldarelli R., Barsi G.,  
 RA Blake J., Botelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Welte C., Whitaker C., Wilming L.,  
 RA Wysshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,  
 RA Hayashizaki Y.,  
 RT "Functional annotation of a full-length mouse cDNA collection.",  
 RL Nature 409:685-690(2001).  
 DR EMBL: AK010578; BAB27041.1; -  
 DR MGD; MGI:1888921; Grost.  
 SQ SEQUENCE 742 AA; 84086 MW; CFC4434D96ACEA7 CRC64;

Query Match 25.6%; Score 545; DB 11; Length 742;

Best Local Similarity 34.7%; Pred. No. 2.7e-38;  
 Matches 147; Conservative 64; Mismatches 169; Indels 44; Gaps 13;

QY 5 RGAALATLALCVACALRAGRAQERYSFSPRDELMPLESAYRHADKYSGEHMAESX 64  
 DB 6 RRLAALMA-VAAALALRV-----AAESEPGMDVAAPRLLYAEGTAAYSRDMDPVV 56  
 QY 65 GYLEISLRHLRLDSEAFCHRNCSA---APQ---PEPA---AGIASYPELRLEGL 112  
 DB 57 LNMERALRSRALRLRLRCRTCAETELPMAPDLGLGDPDLSODPGAAALHDLRFEGAV 116  
 QY 113 LRRHCLKRCQGLPAPROSQPSREVLADFORREPYKFLQAPYKANNLPRATAATFEL 172  
 DB 117 LRRACLRRC-LGPPSAHL--SEELDFENKRSPPYVLOVAYFKINKLEAAVAATFEF 173  
 QY 173 LKHDDMMKRNMAVYKSLPGAEYD-IKDLETKSYESLFIKAVAYNGEMRTSITMEL 231  
 DB 174 VGNPEHEMQRNDLYQYTMGSKVKEADFRDLAKPHMHEFRGLYSEKQAEVPHLEA 233  
 QY 232 ALPPEFAFYCLACGSGREIKDFK-----DFYLSIADHYVEVLECKIOCEENT-TPV 284  
 DB 234 ALQGYEVADECCALCGPYDGYDGYNYLDYSADLFOAITDHYVQVLYNCKKCVTELASHP 293  
 QY 285 IGGPYKEFYATMNYLQAFYKYLNDLKNAPCAVSYLLFQNDKVMQOOLVYQYRHDT 344  
 DB 294 SREKPFEDFLPSHNYLQAFAYNIGNYTOATECAKTYLLFPNDVHMQLAYT-----T 348  
 QY 345 WGLSDEH--FQRPPEAVQEFNVTTLQEL---YDFAKENIMDDDEGEVEYVDLLET 397  
 DB 349 AMLGEEBASISIPRENAEEYRRSLERKELLFAYDIFGLPFVDPDSWTPREVIPKRLQ 408  
 QY 398 EETS 401  
 DB 409 KOKS 412

RESULT 7  
 Q9HCB7 PRELIMINARY; PRT; 363 AA.  
 AC Q9HCB7;  
 DT 01-MAR-2001 (TEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)  
 DT 01-MAR-2001 (TEMBLrel. 16, Last annotation update)  
 DE GROS1-S PROTEIN.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC TISSUE-TESTIS;  
 RA MEDLINE-20406537; PubMed-10951563;  
 RX Kaul S.C., Sugihara T., Yoshida A., Nomura H., Madhwa R.;  
 RT "Gros1, a potential growth suppressor on chromosome 1: its identity to  
 RT basement membrane-associated proteoglycan, leprecan."  
 RL Oncogene 19:3576-3583(2000).  
 DR EMBL: AF097431; AAC31018.1; -  
 SQ SEQUENCE 363 AA; 41158 MW; 5B1B6B38679E76A CRC64;

Query Match 25.3%; Score 539; DB 4; Length 363;

Best Local Similarity 35.4%; Pred. No. 3.4e-37;  
 Matches 132; Conservative 59; Mismatches 146; Indels 36; Gaps 10;

QY 6 RGAALATLALCVACALRAGRAQERYSFSPRDELMPLESAYRHADKYSGEHMAESXG 65  
 DB 4 RALKLITTLAVNA--ASQAEVSESGMGWVTDDL-----FADGTAAYARGDMPGVYL 56  
 QY 66 YLEISLRHLRLDSEAFCHRNCSA-----APQPEPAAGIASYPELRLEGLLR 115  
 DB 57 SMERALRSRALRLRLRCRTCAETELPMAPDLGDPDLSODPGAAALHDLRFEGAV 116  
 QY 116 AHCLKRCQGLPAPROSQPSREVLADFORREPYKFLQAPYKANNLPRATAATFELKH 175  
 DB 117 AACLRRC-LGPPAASL--SEEMLEFRKRSPYVLOVAYFKINKLEAAVAATFEVGN 173  
 QY 176 PDDMMKRNMAVYKSLPGAEYD-IKDLETKSYESLFIKAVAYNGEMRTSITMELALP 234  
 DB 174 PENHEMQOONDLYQYTMGSKVKEADFKDLQPHMHEFRGLYSEKQAEVPHLEALQ 233  
 QY 235 DFFKAFYECCLACGSGREIKDFK-----DFYLSIADHYVEVLECKIOCEENT-TPVIGG 287  
 DB 234 EYFAVAECCALCGPYDGYDGYNYLDYFNADLFOAITDHYIQLVLYNCKKCVTELASHPSRE 293  
 QY 288 YPVKPFATMNYLQAFYKYLNDLKNAPCAVSYLLFQNDKVMQOOLVYQYRHDTWGL 347  
 DB 294 KPFPEDFLPSHNYLQAFAYNIGNYTOAGECAKTYLLFPNDVHMQLAYT-----L 348  
 QY 348 SDEHQ--PRPE 357  
 DB 349 GEEHTRISIGPREQ 361

RESULT 8  
 Q9RLJ8 PRELIMINARY; PRT; 728 AA.  
 AC Q9RLJ8;  
 DT 01-MAY-2000 (TEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TEMBLrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)  
 DE LEPRECAN.  
 GN LEPREL.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-99386987; PubMed-10455179;  
 RA Massenbove-McCarthy D.J., McCarthy K.J.;  
 RT "Molecular characterization of a novel basement membrane-associated  
 RT proteoglycan, leprecan."  
 RL J. Biol. Chem. 274:25004-25017(1999).  
 DR EMBL: AF087433; AAD51875.1; -  
 DR InterPro: IPR000886; ER target.  
 DR PROSITE: PS00014; ER\_TARGET; UNKNOWN 1  
 SQ SEQUENCE 728 AA; 82389 MW; 06AFB6972BF3EE1F CRC64;

Query Match 25.1%; Score 534.5; DB 11; Length 728;  
 Best Local Similarity 35.2%; Pred. No. 2.1e-37;  
 Matches 146; Conservative 60; Mismatches 168; Indels 41; Gaps 13;



RESULT 11  
 O9NA33 PRELIMINARY; PRT: 476 AA.  
 AC O9NA33:  
 DT 01-OCT-2000 (TREMblrel. 15, Created)  
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
 DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)  
 DE Y3JF8A.26 PROTEIN.  
 GN Y3JF8A.26.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 ON NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Matthews L.;  
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=99069613; PubMed=9851916;  
 RA none;  
 RT "Genome sequence of the nematode C. elegans: A platform for  
 investigating biology."  
 RL Science 282:2012-2018(1998).  
 DR EMBL; ALJ32862; CAB7032.1;  
 SQ SEQUENCE 476 AA; 56058 MW; 7963305749D03C7D CRC64;

Query Match 17.5%; Score 372; DB 5; Length 476;  
 Best Local Similarity 25.6%; Pred. No. 1,1e-23;  
 Matches 93; Conservative 79; Mismatches 167; Indels 30; Gaps 9;

OY 35 SFPRLDELPLSAVRAHLDKYSGEHMAESXGYLEISRLHRLRDSFAFCHRNCSAPOP 94  
 D 15 SLCDPKVTFEDYDFQFGNENYTERNMPCVAFMKRAIDDFQYDDVYSCRKCNK--QI 72  
 OY 95 EPAAGLASYPELRFLGGLLRACLCRCQGLPAFRSOPRSREVLADFORREPKYELQFA 154  
 D 73 KPAAT--SSSPKIAKFHETSETALCLILRCRKMDFGHQSIKMSVYHDEEKRPYQYMHIC 130  
 OY 155 YFKANNLPKALAAATFLKHPDDEMKRNAAVYKSLFG--ADYTKDLETSYESLFTRA 213  
 D 131 YHOGELAMAVQSAATFLVIANPDDKDIKQSLNWTNMRGYSDDMLIDERKDHEKFTNG 190  
 OY 214 VRAYNGENMRTSTDMELALPDEFKAFYECLAAC-----EGSREIKDFKFLYSTIA 264  
 D 191 VEAYDOQDMGRCVNEFESEALSIIDCKRCLCQDKIDMSVVGNGPEID-----ILLA 244  
 OY 265 DHYVELECKIQCEENLTPVIGYPRVEKFAVTMTHTYLOFAFYKLNLDKNAAPCAVSYLLF 324  
 D 245 SMRASVIRCEHNCYKRLARINGHY-VGHLFAHFEYLFCHFKMKOAGAQAQVAVNYLLF 303  
 OY 325 DQNDKVMQOONLVY--GYHRDTWGLSDHEFPQREAVOFFVNTTLQKELYFAKENIMDD 382  
 D 304 D-DSPIMRNKRYFGKQYK-----NELTFPSQEVLDIYRRRLERAFLEFMKRYVVK 356  
 OY 383 DEGEVEVEYVD 393  
 D 357 DGELEPPEQADD 367  
 RESULT 12  
 O9NV12 PRELIMINARY; PRT: 527 AA.  
 AC O9NV12:  
 DT 01-OCT-2000 (TREMblrel. 15, Created)  
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)  
 DE CNNA FLJ10718 FIS, CLONE NT2RP3001096, WEAKLY SIMILAR TO RATTUS  
 DE NORBETICUS LEPRECAN MRNA (HYPOTHETICAL 60.4 KDA PROTEIN).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Itoigai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
 RA Nishikawa T., Nagai K., Sugano S., Shitatori A., Sudo H.,  
 RA Magatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,  
 RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,  
 RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,  
 RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuno Y.,  
 RA Ninomiya K., Iwayanagi T.,  
 RT "NEDO human cDNA sequencing project."  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=LUNG CARCINOMA;  
 RA Strausberg R.;  
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AK001580; BAA91769.1;  
 DR EMBL; BC005029; AAH05029.1;  
 DR InterPro; IPR000866; ER\_target.  
 DR PROSITE; PS00014; ER\_TARGET; UNKNOWN\_1.  
 KW Hypothetical protein  
 SQ SEQUENCE 527 AA; 60386 MW; D3B80AE780F6D8DD CRC64;

Query Match 12.1%; Score 258.5; DB 4; Length 527;  
 Best Local Similarity 29.7%; Pred. No. 7,1e-14;  
 Matches 63; Conservative 40; Mismatches 90; Indels 19; Gaps 5;

OY 181 MKRNAAVYKSLPGAED-YIKDLETKEYESLFIKRAVAYNGENMRTSTDMELALPDEFKA 239  
 D 3 MOONINENRATAGVEALQLYDREAKPHMESYNACVKEADDFEALIHFEQALREYFEVE 62  
 OY 240 FYECLACGSGREIKDEF-----DEYLSIDHYVEVELECKIQCEENLTPVIGY-PRK 292  
 D 63 DTECRTECGEPQREPEYELGYKAGLIEALADHYMOVLVCOHECVRELATPRGLSPLEN 122  
 OY 293 FVATMYHTLPAYKYLNDKNAAPCAVSYLLFDQNDKVMQOONLVYGYHRDTWGLSDHEF 352  
 D 123 FLPLHYDYLQFAYRYRGVEYKALCAKAYLLCHPDDDEVDLDNVYDE-----SLDDSI 176  
 OY 353 QP-----RPEAVOFFVNTTLQKELYDFAKENI 379  
 D 177 DPASIEAREDLTFEVRKRIESELKSAEGL 208  
 RESULT 13  
 O88836 PRELIMINARY; PRT: 545 AA.  
 AC O88836:  
 DT 01-NOV-1998 (TREMblrel. 08, Created)  
 DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE CHROMOSOME 6 BAC-284H12 (RESEARCH GENETICS MOUSE BAC LIBRARY) COMPLETE  
 DE SEQUENCE (RESEARCH GENETICS MOUSE BAC LIBRARY) (GENE RICH CLUSTER, B  
 DE GENE).  
 DE B.  
 GN B.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98112780; PubMed=9445485;  
 RA Ansari-Lari M.A., Oeljen J.C., Schwartz S., Zhang Z., Muzny D.M.,  
 RA Lu J., Gorrell J.H., Chinnault A.C., Belmont J.W., Miller W.,  
 RA Gibbs R.A.;  
 RT "Comparative sequence analysis of a gene-rich cluster at human  
 RT chromosome 12p13 and its syntenic region in mouse chromosome 6.";  
 RL Genome Res. 8:29-40(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.







GenCore version 4.5  
Copyright (c) 1993 - 2000 Comugen Ltd.

OM protein - protein search, using sw model

Run on: June 8, 2002, 00:10:15 ; Search time 32.49 Seconds

(Without alignments)  
301.467 Million cell updates/sec

Title: US-09-729-674-2

Perfect score: 2130

Sequence: 1 MEGRGGAALLALCVACA.....DDEGEVEVDLLEFETS 401

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_AA:\*  
1: /cgn2\_6/prodata/2/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/2/1aa/5A\_COMB.pep:\*  
3: /cgn2\_6/prodata/2/1aa/5A\_COMB.pep:\*  
4: /cgn2\_6/prodata/2/1aa/5A\_COMB.pep:\*  
5: /cgn2\_6/prodata/2/1aa/5A\_COMB.pep:\*  
6: /cgn2\_6/prodata/2/1aa/5A\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	94.5	4.4	816	2	US-08-785-310A-8
2	94.5	4.4	816	2	US-08-816-693A-53
3	94.5	4.4	816	2	US-08-885-291-53
4	94.5	4.4	816	4	US-09-496-672-53
5	92.5	4.3	806	1	US-08-451-715A-6
6	91.5	4.3	486	1	US-08-328-256-12
7	90	4.2	531	2	US-08-923-536A-12
8	89.5	4.2	855	2	US-08-816-693A-2
9	89.5	4.2	855	3	US-08-885-291-2
10	89.5	4.2	855	3	US-09-496-672-2
11	87	4.1	548	1	US-08-333-358-2
12	87	4.1	548	1	US-08-463-694-2
13	87	4.1	548	1	US-08-694-501-2
14	87	4.1	556	4	US-09-106-194-3
15	87	4.1	1708	1	US-08-493-092-2
16	87	4.1	1708	1	US-08-508-836A-2
17	87	4.1	3056	1	US-08-508-836A-8
18	87	4.1	3056	2	US-08-629-001A-3
19	87	4.1	3056	2	US-08-874-266-2
20	87	4.1	3056	4	US-08-642-274D-3
21	87	4.1	3056	4	US-08-952-127-3
22	87	4.1	3056	4	US-08-952-014C-3
23	86	4.0	590	2	US-08-980-060-2
24	86	4.0	846	3	US-09-307-185-2
25	85.5	4.0	846	3	US-08-885-291-55
26	85.5	4.0	846	3	US-09-107-847-2
27	85.5	4.0	846	3	US-09-496-672-55

28	84.5	4.0	1513	5	PCT-US93-03076-2	Sequence 2, Appl
29	84	3.9	557	1	US-08-328-256-10	Sequence 10, Appl
30	84	3.9	747	2	US-08-816-693A-51	Sequence 51, Appl
31	84	3.9	747	3	US-08-885-291-51	Sequence 51, Appl
32	84	3.9	747	4	US-09-496-672-51	Sequence 14, Appl
33	83.5	3.9	399	4	US-09-347-801-14	Sequence 20, Appl
34	83.5	3.9	405	4	US-09-347-801-20	Sequence 2, Appl
35	83.5	3.9	688	2	US-09-016-000-2	Sequence 11, Appl
36	83	3.9	434	1	US-08-328-256-11	Sequence 2, Appl
37	83	3.9	436	2	US-08-307-588-2	Sequence 2, Appl
38	83	3.9	557	1	US-08-471-454-2	Sequence 2, Appl
39	83	3.9	557	2	US-08-466-974-2	Sequence 2, Appl
40	83	3.9	557	2	US-08-471-453-2	Sequence 4, Appl
41	83	3.9	557	2	US-08-307-588-4	Sequence 4, Appl
42	83	3.9	625	2	US-08-980-060-4	Sequence 4, Appl
43	83	3.9	625	4	US-09-307-185-4	Sequence 4, Appl
44	83	3.9	824	2	US-08-785-310A-7	Sequence 7, Appl
45	82.5	3.9	608	4	US-09-315-793-22	Sequence 22, Appl

#### ALIGNMENTS

RESULT 1

US-08-785-310A-8

; Sequence 8, Application US/08785310A

; Patent No. 5840532

; GENERAL INFORMATION:

; APPLICANT: McKnight, Steven L.

; TITLE OF INVENTION: Neuronal PAS Domain Protein

; NUMBER OF SEQUENCES: 8

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP

; STREET: 268 BUSH STREET, SUITE 3200

; CITY: SAN FRANCISCO

; STATE: CALIFORNIA

; COUNTRY: USA

; ZIP: 94104

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: IBM PC compatible

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/785,310A

; FILING DATE: 21-JAN-1997

; CLASSIFICATION: 536

; ATTORNEY/AGENT INFORMATION:

; NAME: OSMAN, RICHARD A.

; REGISTRATION NUMBER: 36,627

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 343-4341

; TELEFAX: (415) 343-4342

; INFORMATION FOR SEQ ID NO: 8:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 816 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

; US-08-785-310A-8

Query Match 4.4%; Score 94.5; DB 2; Length 816;  
Best local Similarity 19.8%; Pred. No. 0.46;  
Matches 62; Conservative 41; Mismatches 123; Indels 87; Gaps 14;

QY 126 LPPEFOSQPREVLADQREPKFLQFAVFKANNLPKALIAAHTFLKIPDDMK--R 183  
DB 120 LPADVMDNLNLPPEQSEVVKIL-----SSHMLVTQSPSPPEFLKSDN 164



```

: Sequence 53 Application US/09496672
: Patent No. 6291429
: GENERAL INFORMATION:
: APPLICANT: Takahashi, Joseph S.
: APPLICANT: Turek, Fred W.
: TITLE OF INVENTION: CLOCK GENE AND GENE PRODUCT
: FILE REFERENCE: 0290-5
: CURRENT APPLICATION NUMBER: US/09/496,672
: CURRENT FILING DATE: 2000-02-03
: PRIOR APPLICATION NUMBER: 08/885,291
: PRIOR FILING DATE: 1997-06-30
: PRIOR APPLICATION NUMBER: 08/816,693
: PRIOR FILING DATE: 1997-03-13
: NUMBER OF SEQ ID NOS: 55
: SOFTWARE: Patent In Ver. 2.0
: SEQ ID NO 53
: LENGTH: 816
: TYPE: PRT
: ORGANISM: Mus musculus
: US-09-496-672-53

```

Query Match	4.48;	Score 94.5;	DB 4;	Length 816;
Best Local Similarity	19.88;	Pred. NO. 0.46;		
Matches 62;	Conservative 41;	Mismatches 123;	Indels 87;	Gaps 14

[illegible]

RESULT 5  
US-08-451-715A-6

Sequence 6 Application US/08451712A  
Patent No. 5801013  
GENERAL INFORMATION:  
APPLICANT: Tao, Jianshi  
APPLICANT: Qul, Yan  
APPLICANT: Houman, Fariba  
APPLICANT: Shen, Xiaoyu  
APPLICANT: Schimmel, Paul R.  
TITLE OF INVENTION: Helicobacter Aminoacyl-tRNA Synthetase  
TITLE OF INVENTION: Proteins, Nucleic Acids and Strains Comprising Same  
NUMBER OF SEQUENCES: 67  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Millitia Drive  
CITY: Lexington  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02173  
COMPUTER READABLE FORM:

```

1      MEDIUM TYPE: Floppy disk
2      COMPUTER: IBM PC compatible
3      OPERATING SYSTEM: PC-DOS/MS-DOS
4      SOFTWARE: PatentIn Release #1.0, Version #1.30
5      CURRENT APPLICATION DATA:
6      APPLICATION NUMBER: US/08/451,715A
7      FILING DATE: 26-MAY-1995
8      CLASSIFICATION: 435
9      ATTORNEY/AGENT INFORMATION:
10     NAME: Brook, David E.
11     REGISTRATION NUMBER: 22,592
12     REFERENCE/DOCKET NUMBER: CPI94-25
13     TELECOMMUNICATION INFORMATION:
14     TELEPHONE: 617-861-6240
15     TELEFAX: 617-861-9540
16     INFORMATION FOR SEQ ID NO: 6:
17     SEQUENCE CHARACTERISTICS:
18     LENGTH: 806 amino acids
19     TYPE: amino acid
20     TOPOLOGY: linear
21     MOLECULE TYPE: protein
22     OS-08-451-715A-6

```

Query Match	4.3%;	Score 92.5;	DB 1;	Length 806;
Best Local Similarity	18.6%;	Pred. No. 0.73;		
Matches 57;	Conservative 41;	Mismatches 95;	Indels 113;	Gaps 12;

OY	177	DDEAKNNM-AYVSLG-AEDYKJLDETSYESLFRAVRANGEMPRISIDME----	230
	178	I:::K::: :	
Db	176	DTEVOKELVOYIKITINVAELLKDLETLE-----NIMPSQVILTOKNM1	221
OY	231	--LALPDFEKAECIAACEGSREIKDFD-----FYSIADHY-----VEV	270
	232	: :	
Db	222	GKSIGLOFGRIADECKACNGIDEIEFTTRADTIVGYIAIAEPHPLVEHAIRVSQ	281
OY	271	LECKI-----OCEENLPFVIGGYVEAFV-----	284
	272	: :	
Db	282	EDSKITAILNTTOREALEKKCAFGLVYAIPHILTKOIPWVANPALATYSGALMGVP	341
OY	295	-----ATMHY-----LOFAYYKLNDKNAPCAVSYLEFDONDKVMCOONLY	336
	296	: :	
Db	342	ACDERDEFEEANLYHIPIKVITOSPONDPHRKEEVLKXSGEMS-----DLSSSVARBOII	395
OY	337	YY-----OYHDRTGLSDSEHQPPPEAOFNPVTTL--OKELYFAKENIMDDO	383
	338	: :	
Db	396	AIFPEKNLGKRVIYIRLODMGVSNQRUWGAPDIMIHONGCIYPERQULPTLREDVIDOS	455
OY	384	EGEVYE	389
	385	: :	
Db	456	EGNPLE	461

## RESULT 6

US-08-328-256-12  
Sequence 12, Application US/08328256  
Patent No. 5643749  
GENERAL INFORMATION:  
APPLICANT: REVEL, Michel  
APPLICANT: ABRAMOVICH, Carolina  
APPLICANT: RATOVIRSKI, Edward  
TITLE OF INVENTION: SOLUBLE INTERFERON ALPHA  
TITLE OF INVENTION: PREPARATION AND USE  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROWDY AND NEIMARK  
STREET: 419 Seventh Street, N.W., Suite 3000  
City: Washington  
STATE: D.C  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

Query Match	4.3%	Score 91.5	DB 1	Length 496
Best Local Similarity	20.3%	Pred. No. 0.44		
Matches 84	Conservative 62	Mismatches 157	Indels 111	Gaps 22

RESULT 7  
US-08-923-536A-12  
: Sequence 12, Application US/08923536A

APPLICANT: SAKAI, Yasuyoshi  
APPLICANT: KATO, No. 59654260  
APPLICANT: SHIBANO, Yuji  
TITLE OF INVENTION: PROTEIN DISULFIDE ISOMERASE GENE DERIVED  
TITLE OF INVENTION: FROM STRAIN OF METHYLOTROPIC YEAST  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.

Query Match	4.2%;	Score 90;	DB 2;	length 531;
Best Local Similarity	19.3%;	Pred. No. 0.7;		
Matches 79;	Conservative	67;	Mismatches 183;	Indels 80;
				Gaps 16;

RESULT 8  
US-08-816-693A-2  
: Sequence 2, Application US/08816693A

```

1  GENERAL INFORMATION:
2
3  APPLICANT: Takahashi, Joseph S
4
5  APPLICANT: Turek, Fred W
6
7

```

```

Query Match: 4.2%; Score 89.5; DB 4; Length 855;
Best Local Similarity: 20.3%; Pred. No. 1.7; Indels 81; Gaps
Matches 58; Conservative 36; Mismatches 111;
QY 110 GGLIRRAHCLKCKOGILPAFRQSPKREVLADFORREPKYFIQAFYKANKNL-PKAIAAA 168
      | : : : : : | : : : : : | : : : : : |
Db 129 GSIIYSESVTSLLEHLPSDLVQSOJFNFIPGEHSESVYKILSTHLESLSPLEPYLKS 188
      | : : : : : | : : : : : | : : : : : |
QY 169 H-----TFLLKHPDDEMMKRRMAYKSKP-----GAEDYIKDLFTKSYES- 208
      | : : : : : | : : : : : | : : : : : |
Db 189 NQLEFCCHMLRGITIDPREPSTSEYVRIQGFKSLTSTSTGNGFEGTIDRTIRPSTEDR 248
      | : : : : : | : : : : : | : : : : : |
QY 209 -LPIRAVRAVNGENWRTSTIDMELALPDFKAFYECLACE--GSREIKDFKDFYSLIAD 265
      | : : : : : | : : : : : | : : : : : |

```

Db 249 VCFATVR-----LATPOFIKEMCTVEEBNEEFSTRHSLEMKFLF---D 290  
Qy 266 HVEVELECKIOCEENTLTVIGGYVEKEFVATWYHYLOFAYKRLNDLKNAPCAVSYLLFD 325  
Db 291 H-----RAPPILIGLPEVEVIGTSGYD-----YIVVDLEMLAKHEHLMQYG 332  
Qy 326 ONDK-----VMOONLVYQYHRDWTGSLDEHPORPEAV 359  
Db 333 GKSCYVRFELTKGOQWIMLQTHYITVHQ-----WNSRPEFI 369  
  
RESULT 11  
US-08-333-358-2  
; Sequence 2, Application US/08333358  
; Patent No. 5571696  
; GENERAL INFORMATION:  
; APPLICANT: EVANS Ph.D., RONALD M.  
; APPLICANT: MANGELSDORF Ph.D., DAVID J.  
; APPLICANT: ONG MS., ESTELITA S.  
; APPLICANT: ORO Ph.D., ANTHONY E.  
; APPLICANT: BORMEYER Ph.D., UWE K.  
; APPLICANT: GIGIERE Ph.D., VINCENT NMN  
; APPLICANT: YAO Mr., TSO-PANG NMN  
; TITLE OF INVENTION: NOVEL RECEPTORS  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark  
; STREET: 444 So. Flower St., Suite 2000  
; CITY: Los Angeles  
; STATE: CA  
; COUNTRY: US  
; ZIP: 90071-2921  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/333,358  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/761,068  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Reiter Ph.D., Stephen E.  
; REGISTRATION NUMBER: 31192  
; REFERENCE/DOCKET NUMBER: P31 8936  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 535-9001  
; TELEFAX: (619) 535-8949  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 548 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-333-358-2  
  
Query Match 4.1%; Score 87; DB 1; Length 548;  
Best Local Similarity 20.2%; Pred. No. 1.5;  
Matches 93; Conservative 48; Mismatches 156; Indels 164; Gaps 21;

Db 271 DYTASGFEPPYCSFTNGETSPVTSMAELEHLAONTSKSH-LETCOYLKEELQOITWOTFL 329  
Qy 126 REVLADEFORREPKFLOPAFVFRANLPRKIAAAHTFLFKHPDDEMKRNMAVYKSLPG-- 193  
Db 330 OEIENYONKOREVWMOJCAKIT-----EAIQVVEFARKIDGFM 370  
Qy 194 ---AEDYIKDLETKSYESLFIKRAVAYNGENMRSTITMELALPDFFKAFYECJLAACG- 249  
Db 371 ELQONDQIVLLKAGSLLEVFIKRCRAFDSON-NTVYFGKAYSPDVFKSL-----GCEDF 424  
Qy 250 -----SREIKDEFKDFLSTADH-----YVEVELECKIOCEENTLTVI 285  
Db 425 ISFVEFEKSLCSMHLTEDEIALFSAFVLMGADRSMLOEKVKIEKLOKIO----- 475  
Qy 286 GGYVEKEFVATWYHYLOFAYKRLNDLKNAPCAVSYLLFDQNDKWOONLVYQYHR 342  
Db 476 -----LALQHVLO-KNHRDQILTKLICKYSTLALGHRHEKMAKRAIYPDIVR 525  
Qy 343 DTWGISDEHFOPRPEAVQFENVYTLQKELYDPAKENIMDD 383  
Db 526 -----LHFPF-----LYKELTSEFEPAQID 547

RESULT 12  
US-08-463-694-2  
; Sequence 2, Application US/08463694  
; Patent No. 5696233  
; GENERAL INFORMATION:  
; APPLICANT: EVANS Ph.D., RONALD M.  
; APPLICANT: MANGELSDORF Ph.D., DAVID J.  
; APPLICANT: ONG MS., ESTELITA S.  
; APPLICANT: ORO Ph.D., ANTHONY E.  
; APPLICANT: BORMEYER Ph.D., UWE K.  
; APPLICANT: GIGIERE Ph.D., VINCENT NMN  
; APPLICANT: YAO Mr., TSO-PANG NMN  
; TITLE OF INVENTION: NOVEL RECEPTORS  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark  
; STREET: 444 So. Flower St., Suite 2000  
; CITY: Los Angeles  
; STATE: CA  
; COUNTRY: US  
; ZIP: 90071-2921  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/463,694  
; FILING DATE: 05-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/761,068  
; FILING DATE: 17-SEP-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Reiter Ph.D., Stephen E.  
; REGISTRATION NUMBER: 31192  
; REFERENCE/DOCKET NUMBER: P31 8936  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 535-9001  
; TELEFAX: (619) 535-8949  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 548 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-463-694-2



[illegible]

```

COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/106,194
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Leitch, Debra K
REGISTRATION NUMBER: 32,619
REFERENCE/DOCKET NUMBER: 96-11
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-442-6674
TELEFAX: 206-442-6678
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 556 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-106-194-3

```

```

Query Match 4.1% Score 87; DB 4; Length 556;
Best Local Similarity 20.2% Pred. No. 1.6;
Matches 93; Conservative 48; Mismatches 156; Indels 164; Gaps 21;

```

```

QY 16 CVAC-----ALRAGR-----AQYRSFSSFPD-----ELMPLESAYRHA--- 51
D 166 CLAVGMSDAVKFGMSKMKORDSLYAEVQKHMQQQDHOQPEAEPLPTPTYSANG 225
QY 52 -----LDKSGEHWAE-----SKGYLEISRLHRLRDSFAFCHRNCSAAPQ-- 94
D 226 LTELHDDLSNYIDGHTPGSKADSAVSSFYLDIOPSPDQSGIDINGI-----KPEPIC 278
QY 95 --EPAGLASV-----PELRFGGLLRRAHCLKRC--KQGLPAFROSPS 135
D 279 DYTASGFFPCSTNGSTPTVSMALDELHLAONISKSH-LETQOYLREELQOITWQTF 337
QY 136 REVLADEFRREPKYFLQAFYKANNLPKALAAHTFLKHDPDEMKRNMAVYKSLPG-- 193
D 338 QEEIENYQNKQREVMWQCAIKIT-----EAIQYVVEFAKRIDQFM 378
QY 194 ---AEDYTKDLETYSYSLFTIRAVRAYNGENWRTSTIDMELALPDFEKAFTYECIAACG- 249
D 379 ELCONDQIVLLKAGSLLEVYFTRMCRAPDSQ--NTYVEDGKYASPDVEKSL-----GCEDF 432
QY 250 -----SREIKDKDFYLSIADH-----YVEVLDECKICQCEENTPTVI 285
D 433 ISEVFERGKSLCSMHLTEDELTALSAFVLMGSDNSWLOEKVKIEKLOOKID----- 483
QY 286 GGYEVEKVFATMYHLOFAYYKLLNDLKNAPCAVS---YLLEPDQNDKWOONLVYYQYHR 342
D 484 -----LALQHVLO-KNHRDEGILTKLCKYSTLRALGCRHTEKLMARKAITYPDLYR 533
QY 343 DTWGLSDEHQPREEAVQFFNVTTLOKELYDFAKENIMDDDD 383
D 534 -----LHFPF-----LYKELPTSEFEPAQID 555

```

```

RESULT 15
US-08-493-092-2
Sequence 2, Application US/08493092
Patent No. 5728807
GENERAL INFORMATION:
APPLICANT: Shioh, Josef
APPLICANT: Taglie, Danilo A.
APPLICANT: Collins, Francis S.

```

```

TITLE OF INVENTION: Ataxia-Telangiectasia Gene
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Reising, Ethelington, Barnard & Perry
STREET: P.O. Box 4390
CITY: Troy
STATE: Michigan
COUNTRY: US
ZIP: 48099
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/493,092
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Kohn, Kenneth I.
REGISTRATION NUMBER: 30,955
REFERENCE/DOCKET NUMBER: P-310 (TAV)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (810) 689-3500
TELEFAX: (810) 689-4071
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1708 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-493-092-2

```

```

Query Match 4.1% Score 87; DB 1; Length 1708;
Best Local Similarity 17.9% Pred. No. 8.8;
Matches 77; Conservative 56; Mismatches 135; Indels 160; Gaps 16;

```

```

QY 84 CHRNCSAQPPEPAGLASYPRLFGGLRRACLKCKOGLPAPFROSPSREVLADFQ 143
D 525 CLRHFSQTSRSTTPANDSESE-----HFFRCIDP-----KKSQRTMLAVVDYM 568
QY 144 RREPYK-----FLQAFYKANNLPKALAAHTFL-----KHDPDEMKRNMA 186
D 569 RROKRPSSTIIFNDAFMDLVLYEVAKYAOCSAHPFALLAETIYADKKSMDDCKSLA 628
QY 187 Y-----YKSLPGAEDYIKDLETKSY-----ESLF-----IRAVRAY 217
D 629 FEESQSTTSSLSKSEKSEETGISLDLLLEIYNSIGEPDSSLVCGCGGKMLQITRLKTY 688
QY 218 NGE-NWRTSIT--DMELALPD-----FVK----- 238
D 689 EHEAMGKALTYTDELTAIPSTROAGIQAOLNGLCHILSYVLKGLDYENKMCPELE 748
QY 239 -----AFYECLAACESREIKDKDFYLSIADHYVEYLE 772
D 749 ELHYQAMRNQMOMCHTSVSKEVEGTSYHESLYNALQSLRDRERSTYESL--KYARVKE 806
QY 273 CKICEENLTPTVIGGYEVEKVFATMYHLOFAYKYKLDLKNAPCAVSYLLFQNDKVMQ 332
D 807 VEEMCKRSLSESYSLYPT-----LSRLQAIIGELSIGELFSRSHVTHRQ 849
QY 333 QNLVYYQYHRTWGLSDEH--OPREAVQFFNVTTLOKELYDFAKENIMDDDEGEVVE 389
D 850 LSEVYIKWQKHSQLLKSDSPSFQEP-----TALRTVYILEITMEKEDNSQRCIK 900
QY 390 YV--DDELEL 397
D 901 DILTKHLEVEL 910

```

Search completed: June 8, 2002, 02:08:49

Job time: 714 sec

---

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 8, 2002, 00:14:20 ; Search time 49.19 Seconds

(without alignments)  
783.326 Million cell updates/sec

Title: US-09-729-674-2

Perfect score: 2130  
Sequence: 1 MEPCRGCAALALLLCVACA.....DDEGEVEVEYVDLLEETS 401

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : PIR\_71:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1019	47.8	431	2	A56822 synaptonemal compl
2	106	5.0	1345	2	A64430 DNA-directed RNA p
3	104	4.9	317	2	B95113 competence protein
4	104	4.9	317	2	A97982 competence protein
5	104	4.9	853	2	H70470 hypothetical prote
6	98.5	4.6	1575	2	G83905 conserved hypothet
7	98	4.6	376	2	T32009 hypothetical prote
8	98	4.6	484	2	E70031 hypothetical prote
9	97.5	4.6	910	2	T41331 proline dehydrogen
10	97.5	4.6	1162	2	C81297 hypothetical prote
11	96	4.5	788	2	A71076 conserved hypothet
12	95.5	4.5	240	2	E90139 conserved hypothet
13	95	4.5	412	2	H65011 probable transamin
14	94.5	4.4	292	2	S64361 hypothetical prote
15	94	4.4	435	2	B81279 probable fibronect
16	94	4.4	922	2	S37871 hypothetical prote
17	93	4.4	422	2	AD2191 hypothetical prote
18	93	4.4	433	2	D91111 probable adherence
19	93	4.4	433	2	H85956 probable cytooxin
20	93	4.4	434	2	B97018 xre family DNA-bin
21	93	4.4	1101	2	AB2481 hypothetical prote
22	92.5	4.3	729	2	E81410 transcription regu
23	92	4.3	411	2	E83056 transaminase (EC 2
24	92	4.3	534	2	T41081 hypothetical prote
25	92	4.3	1132	2	T44001 major DNA binding
26	91.5	4.3	431	2	C86812 conserved hypothet
27	91.5	4.3	684	2	T33785 hypothetical prote
28	91.5	4.3	686	1	ALDYAT amyliase A (EC 3.2.
29	91.5	4.3	870	2	T01310 hypothetical prote

30	91.5	4.3	1416	2	T20823 hypothetical prote
31	90.5	4.2	244	2	T19750 hypothetical prote
32	90.5	4.2	403	1	Q08A4E hypothetical prote
33	90.5	4.2	588	2	T37542 conserved hypothet
34	90.5	4.2	625	2	C75044 threonyl-tRNA synt
35	90.5	4.2	632	2	AE2382 acetylhydroxy acid
36	90	4.2	1112	2	T47784 hypothetical prote
37	89.5	4.2	199	2	AD1230 phosphoglycerate m
38	89.5	4.2	376	2	B84277 hypothetical prote
39	89.5	4.2	397	2	T23642 hypothetical prote
40	89.5	4.2	407	2	A83279 hypothetical prote
41	89.5	4.2	416	2	S16658 mobilization prote
42	89.5	4.2	631	1	BHTLA hemocyanin chain a
43	89	4.2	217	2	H71507 hypothetical prote
44	89	4.2	333	2	C84792 hypothetical prote
45	89	4.2	335	2	B71485 probable tpr repea

#### ALIGNMENTS

RESULT 1  
A56822  
synaptonemal complex protein SC56 - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 05-Jan-1996 #sequence\_revision 05-Jan-1996 #text\_change 04-Mar-2000  
C:Accession: A56822; S20742  
R:Chen, Q.; Pearlman, R.E.; Moens, P.B.  
Biochem. Cell Biol. 70, 1030-1038, 1992  
A>Title: Isolation and characterization of a cDNA encoding a synaptonemal complex pr  
A:Reference number: A56822; M0ID:93213429  
A:Accession: A56822  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-431 <CH>  
A:Cross-references: EMBL:X65454; NID:957191; PIDN:CAA46449.1; PID:957192  
A:Experimental source: testis  
A>Note: sequence extracted from NCBI backbone (NCBIP:129004)  
C:superfamily: rat synaptonemal complex protein SC56

Query Match	47.8%	Score 1019;	DB 2;	Length 431;
Best Local Similarity	53.6%	Pred. No. 1.5e-75;		
Matches	201;	Conservative	53;	Mismatches 91;
			Indels	30;
			Gaps	5;
QY	42	MPLEBSAYRAHLDKRYSGEMAESKYLEISRLHRLDSEAFCHRNCS--AAPPEPAAG	99	
DB	1	MPLATAYGHALDEQYEGESMRESARYLEAALRLHRLDSEAFCHANSGPATSQPRAPG	60	
QY	100	LASVP-----ELRLFGGLLRRAHCLCKRCOGIPAFROSQSPREVLADPQRE	146	
DB	61	-----PGSDNDEGDEDMARELRRLFGHVLERRACLRCKRTTPAHOVPYPSQLDLDFQNR	116	
QY	147	PYKFLQFAVFKRANLPRKIAAHTFLKHPDDEMKRNNMAYYS-LPGADYIKDLETKS	205	
DB	117	PYGLYHAFKFRANLREKAAVAATYFLQRPKHELTAKLYLVYRMDLIGESLTDLEAOP	176	
QY	206	YESIFRAVRAIVYGENMKRTISITDMELALPPEFAFYCLACBSRSRIKPKFYLEIAD	265	
DB	177	YEAVFLOAVKLTVNSGDFRSSTEHMERALADYMTVFACLAGCBGAHGOVDFKDEYPAIAD	236	
QY	266	HYAEVLECKTIOGENTLPPVIGGYVPEKFAVTMYHYLOFAYYKLNDAKNAAPCAVSYLLFD	325	
DB	237	LPFESLDCKKVCENANLTPNNGGFVDFVATMTHYLOFAYYKLNDAVQAARSASTYLLFD	296	
QY	326	QNDKVMQNLVYQYHBDTWGLSDEHFQPRPEAVQFNVTTLOKELYDPAKINIMDDDEG	385	
DB	297	PKDSVMQNLVYVYRFHRAHGLBEEDFQPREAVLVHNGQSELERLDFTTHMYLQSDDE-	355	
QY	386	EVEVEYVDLLEET 400		
DB	356	-----MELEET 361		

RESULT 2  
A64430  
DNA-directed RNA polymerase (EC 2.7.7.6) subunit A' (intein-containing) - Methanococcus  
N:Contains: Intein  
C:Species: Methanococcus jannaschii  
C>Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 17-Mar-2000  
C:Accession: A64430  
R:Butt, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,  
; Reisch, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;  
rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.  
Science 272, 1058-1073, 1996  
A:Authors: Kane, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C  
A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii  
A:Reference number: A64300; MUID:96337999  
A:Accession: A64430  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-1345 <BUL>  
A:Cross-references: GB:067547; GB:L77117; NID:g1591695; PID:g1591696; TIGR:MJ1042; PID:g  
C:Genetics:  
A:Map position: FOR974597-978634  
A:Start codon: TTG  
C:Superfamily: yeast DNA-directed RNA polymerase III 160K chain  
C:Keywords: nucleotidyltransferase; protein splicing; transcription

Query Match 5.0%; Score 106; DB 2; Length 1345;  
Best Local Similarity 23.8%; Pred. No. 3.3;  
Matches 55; Conservative 34; Mismatches 82; Indels 60; Gaps 12;  
QY 189 KSLGAEYITDLETKSESIFIRAVRYNG-----EMWRISITDMELALDFFFAF 240  
DB 714 KKMGIPIWIK-TAPKTIKKEFL--SAVFSESLTPKIRNHGTSFKLSFKIKIEIF 769  
QY 241 YECIACGSEGRKIDFDFYLSIDHYVELECKIOCEE-MLTPVIGSGPYEKFEVATWYH 299  
DB 770 DE-----DFIKIDIKEMLKEFG-----TELKRVYEGSLRK--DQYKKYVASTYN 814  
QY 300 YLQF---AYKKLNDLNAACAVSYLF-----DONDKYMQONLYYYQYHRDTWGLSD 349  
DB 815 HKEFFGRIGYTYANKKETLARYAYEYLLTKEKYLKDRNIKKLENNTKFTFDK----- 867  
QY 350 EHPQPREAVGFNVTTLQKELYPAKENIMDDGEVVEVVDLLLEET 400  
DB 868 -----FIKCKLKN--GFVKEKIVSIEETK-VDYVYDITITSET 903

RESULT 3  
B95113  
competence protein Coia [imported] - Streptococcus pneumoniae (strain TIGR4)  
C:Species: Streptococcus pneumoniae  
C>Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 03-Aug-2001  
C:Accession: B95113  
R:Nettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Reed, T.D.; Peterson, S.; Heid  
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,  
nson, T.; Hickey, E.K.; Holt, I.E.  
Science 293, 498-506, 2001  
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,  
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.  
A:Reference number: A95000; MUID:21357209; PMID:11463916  
A:Accession: B95113  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-317 <KUR>  
A:Cross-references: GB:AE005672; PIDN:AAK75099.1; PID:g14972453; GSPDB:GN00164; TIGR:SP4  
A:Experimental source: strain TIGR4  
C:Genetics:  
A:Gene: SP0978

Query Match 4.9%; Score 104; DB 2; Length 317;  
Best Local Similarity 22.1%; Pred. No. 0.7;

Matches 61; Conservative 38; Mismatches 119; Indels 58; Gaps 11;  
QY 13 ALLCVACA---LRAGRAQYERYFSRFPDELMPLESAYRHALDKYSGEHWASXGYLE 68  
DB 23 AYTCPCAGGQHLRLRGSPVTRTHFAHKSLLKDCDFFENSPHANKESLYHMLKRETVQ 82  
QY 69 ISLRHLRLRDSFAFCNRN-----CSAAPQ---PEPAAGLAS--YPELRIFGLIRRAH 117  
DB 83 LEYPLSELKQIADYFVGNLALAEVQCSPLPKVLKERSEGRSGGYOVLMLGOKLMLKE 142  
QY 118 CLKRCQGLPAFROS-----QPSREVL-----ADFORREPKYKLOFAFKAN--- 159  
DB 143 RLTRLQCGFLYFSQMGFYWELDKGQVLLKLYLIDDLKGLKHQYKESYGGSLLE 202  
QY 160 --NLP-KAIAAAHFLKHPDEMKNRMAYKSL---PGAEDYIKDLETKSYESLPIR 212  
DB 203 ILRLPYRKOKISHFTVSEDKDICRYIRQOLYYOVLFWMKEQEAAYOKG-----E 251  
QY 213 AVRAYNGENMRTST-----DMELALPDEFKAFY 241  
DB 252 NILTYGLKEWYQIRPIVKGFEQIQLDLSYQHFY 287

RESULT 4  
A97982  
competence protein [imported] - Streptococcus pneumoniae (strain R6)  
C:Species: Streptococcus pneumoniae  
C>Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 22-Oct-2001  
C:Accession: A97982  
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Buretti, S.; Dehoff, B.S.  
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.  
y, P.; Sun, P.M.; Winkler, M.E.  
J. Bacteriol. 183, 5709-5717, 2001  
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Balitz, R.H.; Jaskunas, S.  
A:Title: Genome of the Bacterium Streptococcus pneumoniae strain R6.  
A:Reference number: A97872; MUID:21429245; PMID:11544234  
A:Accession: A97982  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-317 <KUR>  
A:Cross-references: GB:AE007317; PIDN:AAK99685.1; PID:g15458486; GSPDB:GN00174  
C:Genetics:  
A:Gene: coia

Query Match 4.9%; Score 104; DB 2; Length 317;  
Best Local Similarity 22.1%; Pred. No. 0.7;  
Matches 61; Conservative 38; Mismatches 119; Indels 58; Gaps 11;

QY 13 ALLCVACA---LRAGRAQYERYFSRFPDELMPLESAYRHALDKYSGEHWASXGYLE 68  
DB 23 AYTCPCAGGQHLRLRGSPVTRTHFAHKSLLKDCDFFENSPHANKESLYHMLKRETVQ 82  
QY 69 ISLRHLRLRDSFAFCNRN-----CSAAPQ---PEPAAGLAS--YPELRIFGLIRRAH 117  
DB 83 LEYPLSELKQIADYFVGNLALAEVQCSPLPKVLKERSEGRSGGYOVLMLGOKLMLKE 142  
QY 118 CLKRCQGLPAFROS-----QPSREVL-----ADFORREPKYKLOFAFKAN--- 159  
DB 143 RLTRLQCGFLYFSQMGFYWELDKGQVLLKLYLIDDLKGLKHQYKESYGGSLLE 202  
QY 160 --NLP-KAIAAAHFLKHPDEMKNRMAYKSL---PGAEDYIKDLETKSYESLPIR 212  
DB 203 ILRLPYRKOKISHFTVSEDKDICRYIRQOLYYOVLFWMKEQEAAYOKG-----E 251  
QY 213 AVRAYNGENMRTST-----DMELALPDEFKAFY 241  
DB 252 NILTYGLKEWYQIRPIVKGFEQIQLDLSYQHFY 287

RESULT 5  
h70470  
hypothetical protein ag\_1989 - Aquifex aeolicus

C:Species: *Aquifex aeolicus*  
C:Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 05-Nov-1999  
C:Accession: H70470  
V:Dacker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Over  
Nature 392, 353-358, 1998  
A:Title: The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*.  
A:Reference number: A70300; M0ID:9819666  
A:Accession: H70470  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-853 <AAF>  
A:Cross-references: GB:AE000766; NID:g2984216; PIDN:AAC07758.1; PID:g2984229; GB:AE000651  
A:Experimental source: strain VF5  
C:Genetics:  
A:Gene: aq\_1989

Query Match	4.9%;	Score 104;	DB 2;	Length 853;
Best Local Similarity	20.4%;	Pred. No. 2.6;		
Matches	65;	Conservative	43;	Mismatches 80;
				Indels 130;
				Gaps 15;

Qy	129	FROSPFSEVLAADQREPERYKFLQFAFKANNLEKALIAATFLFKHDDMMKRNNA-X	187
Db	223	FIAGRPNIADYLISIKVRRAIKLFNMWG-----RLEAKKILLFEPDKIPIPRDVLY	273
Qy	188	YKSL-----PQAEYIKDLEH-----KS-YESLFIR-----	212
Db	274	YLALIALNEGKTBALDYAVILLETINPELAKSVLSFLKCHSLADPYRKTKGEKWLEY	333
Qy	213	-AVRAYNGEMRTSITDME-----LALPDEFKAFYECLACEGSRETKD-FKD	258
Db	334	AGVAAIYNGATRKALDEFERKAGKAKYAVYALAKIDYEKA-YELL-----KNKKDRRE	386
Qy	259	FY-----LSIADHYVELECKIQCEENLPRVIGGYEKEFVATMYHLYDAUYKLN	309
Db	387	YYRLLEVIYSTDEDEFLKTFLE-----EIAKYP-----	416
Qy	310	DLKNAACVAVSLLEDDQNDKMOONLYUYQYHNRTWGLSDBNH-QRPREAVQFNNVTTLQ	368
Db	417	-----KLKYEYGYVLEFKKNNWLEAEKYEFNNPYKKAADAVYNNAGDYE	457

```
QY      369 KELYDFAKENIMDDDEGE 386
      1 1 : 1 1 : 1
Db      458 KVL-----ELKDDNSYE 470
```

```

RESULT      6
GB2905
C:Species: Ureaplasma urealyticum
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: GB2905
R:Glass, J.I.; Leftkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Cassell, G.H.
  submitted to Genbank, February 2000
A:Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a mild
A:Reference number: A82870
A:Accession: GB2905
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1575 <GLA>
A:Cross-references: GB:AE002129; GB:AF222894; NID:g6899292; PIDN:AAF30726.1; GSPDB:GN001
A:Experimental source: serovar 3; biovar 1
C:Genetics:
A:Gene: UU317
A:Genetic code: SGC3

```

Query Match	4.68;	Score 98.5;	DB 2;	Length 1575;
Best Local Similarity	21.48;	Pred. No. 17;		
Matches 57;	Conservative 50;	Mismatches 76;	Indels 83;	Gaps 14;
176 PDDEMKRMMAYYK-----SLPGAEIDYIKDELTSYESTFTRAVRAVNGE---NMR 223				

Db	815	PENNQLQTLNETYKKKWSCKLINDHSIFSQDLIEGVNELAYDLNLR--NSDPENEYEYF	872
Oy	224	TSITDMELARDPFKAF-----YELCA-----ACBSRELPKPDYLL	261
Db	873	TEFSIVPTLYLKRFSKSLDHOVTELLRYHNHLENKYIKYKLVFKNNQWTLINKDNF	932
Oy	262	SIAD-HAYEVELECKIOCEENLTPVIGIGYERKVFVATWTHYLOFAVYKLLNDKLNAPCAVS	320
Db	933	PIDQNFVSAIYKKPKMEQN-----YDLKRYI-----FKNDLL-----IYD	968
Oy	321	YLLFDQNDKWOONLVLVYUYQHRDTWGLSDENHPQR--PEA-----VOEFNVTLQ	368
Db	969	EYVERNSK---QN-----YHKLTLYTFSSYFARPLKRPDEGRDYLLPLKYFNATIVR	1020
Oy	369	KELYDFAKEN--IMDDDEGEVVEYU	352
Db	1021	KELLENAINNELFKINDENKTLIKIN	1046

```
QY      369 KELYDFAKEN--IMDDDEGEVEVEYVD 392
      ||| : | | : ||| :: : :
Db      1021 KELLENAINNEILFKNDENKIIRKIN 1046
```

RESULT 7  
T32009  
hypochemical protein K10G6.4 - Caenorhabditis elegans  
C.Species: Caenorhabditis elegans  
C.Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C.Accession: T32009  
R.Davidson, S.; Wohlmann, P.; Mullen, G.  
Submitted to the EMBL Data Library, July 1997  
A.Description: The sequence of C. elegans cosmid K10G6.  
A.Reference number: Z21111  
A.Accession: T32009  
A.Status: preliminary: translated from GB/EMBL/DBJ  
A.Molecule type: DNA  
A.Residues: 1-376 <DAV>  
A.Cross-references: EMBL:AF016659; PIDN:AAB6099.1; GSPDB:GN00020; CESP:K10G6.4  
A.Experimental source: strain Bristol NZ; clone K10G6  
C.Genetics:  
A.Gene: CESP:K10G6.4  
A.Map position: 2  
A.Introns: 9/3; 25/3; 98/2; 161/2; 238/1; 310/2; 333/3

Query Match	4.63;	Score 98;	DB 2;	Length 376;
Best Local Similarity	21.28;	Pred. No. 2.7;		
Matches 68;	Conservative 56;	Mismatches 111;	Indels 86;	Gaps 14;

QY	123	KOGJPAROSOPSEEV-ADOFREPRKFOGAFKANNLPKAIAAHTEFLHNPDEKM	181
Dp	93	KRVSREHLEHKIEDLHAELOKKEBLLKSNQ-----ASELSLKNGEJDL	141
QY	182	KRNMYYS-LPGADYIKDL-----TKSTESLIRAVRANGENMTST	227
Dp	142	KRSBAYAKSOSLOEDYVYSKLHTEKIAENQTVVETLIRKNEBELKLENLMEKOR	201
QY	228	---DMELATDPFKAFUECLACEGSREIKDFKDFLSIADHYVELECKIOCEENTLP	283
Dp	202	QEOJDFAROLADSEFTSVRSSEBKCEQDKL-----IEDHSHQLEKD-RLOPSLEK	255
QY	284	VIGGPRVK-----FVATMYLDFAYTKLNDLKNAAPCAVSYLLFGQNDKVMQNL	335
Dp	256	ADRSQOIEEHRERLELEVSKLYEKLO-----KSA-----EDROKVE----	292
QY	336	VYUOYHNDTWLS-----DENQRPREAVQFPNVTLOKELDYDEAKENIMDDGEV---	388
Dp	293	TTKLNDNRLEVLSVTMKKLESPREARF--NVKKLKELRDSEKRRLSKEKNQIATSE	350
QY	389	-----EYVDDLEEE	399
Dp	351	LKSAKLRLERSYRDVAVIKLO	371

RESULT  
E70031

## hypothetical protein yvcd - Bacillus subtilis

C:Species: Bacillus subtilis

C&gt;Date: 05-Dec-1997 #sequence,revision 05-Dec-1997 #text,change 11-Jan-2000

C:Accession: E70031

R:Kunst, F.; Ogawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter

C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch

A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabel, C.; Ferrati, E.

Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizi, A.; Gall

lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.

A:Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinis

A:Authors: Lader, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mue

Y, M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Potte

Riger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon,

A:Authors: Schleich, S.; Schroeter, R.; Scottone, F.; Sekiguchi, J.; Sekowska, A.; Ser

akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,

T.; Wilters, P.; Wipat, A.; Yamamoto, H.; Yaman, K.; Yasumoto, K.; Yata, K.; Yoshida, K

A:Authors: Yoshikawa, H.F.; Zimstein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.

A:Reference number: A69580; MUID:96044033

A:Accession: E70031

A&gt;Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-484 &lt;XUN&gt;

A:Cross-references: GB:299121; GB:AL009126; NID:92635627; PIDN:CAB15486.1; PID:e1186169;

A:Experimental source: strain 168

C:Genetics:

A:Gene: yvcd

C:Superfamily: unassigned tetratricopeptide repeat proteins; tetratricopeptide repeat hc

F:133-186/Domain: tetratricopeptide repeat homology &lt;TM1&gt;

F:187-320/Domain: tetratricopeptide repeat homology &lt;TM2&gt;

F:221-254/Domain: tetratricopeptide repeat homology &lt;TM3&gt;

Query Match 4.6%; Score 98; DB 2; Length 484;

Best Local Similarity 20.8%; Pred. No. 3.8; Mismatches 106; Indels 94; Gaps 13;

Matches 64; Conservative 44; Mismatches 106; Indels 94; Gaps 13;

QY 57 GEHWAESXGYLEISLRHLRLDSEAFCHRNCSAPOPEPAAGLASYPE-LRLFGGLR 114

Db 2 GKHSSEKHNHOIV-----QLLDQGYFFHK-----GLKAVKERNLTKASKLIQ 45

QY 115 RA-----HCLKRCQGLPFRQSPSR-----VLADQRRPYPFLQPAFYKANN-- 160

Db 46 RAVHLEPDESEMQLAVIYSEMGOYESNDLDYIMANLEAEMP-----ECHYKANNFA 101

QY 161 -----LPKAIAAHFLKHPDEMKRMAYYKSLPGAEYIKDLETKYSLSFIRAVRA 216

Db 102 HLGFOAYKKAAYSDADPDGEFAENDSLDLDDGDSIED-SLYDDDELLVKODRA 160

QY 217 YNGENWFTSTDMELALPDEFKAYECLACEGSREIKDFEYLSIADHYEVLECKIQ 276

Db 161 -----KSLLESGLA-----EAVVALE-----EITTEPELM----- 187

QY 277 CEENUTPVIGIPEKRVATMYHILQFAYYKLANDKNAAPCAVSYLLFDO-NDKVMQONL 335

Db 188 -----SAYNNLALAYFSGVYKAKQAYEVLSHNEGMLHALCNLL 228

QY 336 VYYOYHRD 343

Db 229 VFYTYERE 236

RESULT 9 T41331

probable receptor associated protein - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C&gt;Date: 03-Dec-1999 #sequence,revision 03-Dec-1999 #text,change 03-Dec-1999

C:Accession: T41331

R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Murphy, L.; Harris, D.

submitted to the EMBL Data Library, March 1998

A:Reference number: 221988

A:Accession: T41331

A&gt;Status: preliminary; translated from GB/EMBL/DBD

A:Molecule type: DNA

A:Residues: 1-910 &lt;MOO&gt;

A:Cross-references: EMBL:AL022243; PIDN:CAA18287.1; GSPDB:GN00067; SPDB:SPCC364.05

A:Experimental source: strain 972h; cosmid c364

C:Genetics:

A:Gene: SPDB:SPCC364.05

A:Map position: 2

A:Insertions: 26/1; 47/3; 83/1; 145/3

Query Match 4.6%; Score 97.5; DB 2; Length 910;

Best Local Similarity 21.4%; Pred. No. 9.7; Mismatches 136; Indels 151; Gaps 27;

Matches 94; Conservative 56; Mismatches 136; Indels 151; Gaps 27;

QY 44 LESAVRHADKYSG-EHWAESXGYLEISLRHLRLDSEAFCHRNCSAPOPEPAAGLAS 102

Db 381 LOSAV-----LEQISALQYKKQ--GQYENSLSLETSKI-----DERVVISL 420

QY 103 YPELRLEFGGLLRRAHCLKRCQGLPAPROS-----QPSREYLA 140

Db 421 YPD-----LYHSELVIAFOGVISFRKSILSYDDVAVANVKSNELEFRDTQYTORQMI 473

QY 141 DFORREPKFLQ-----FAFYKANNLKAIAAATFLKH--PPD----- 178

Db 474 EITENAVYIMLRYLKKNKNTSISEYLLSSQRDVMAVEHSLLLYLMMDLAAGTKNA 533

QY 179 -EMKRMNAYYKSLPGAEYIKD--LETKSYE--SLFIRAVAYNG--ENRSTST----- 227

Db 534 KELLE-----SGLTGVED-IKDVLQKKEYLLSVLATVYRDHGVLYQTKKILITGYE 586

QY 228 -----DMELALPDEFKAYECLA-----ACGSREIRDKDFYLSIADHY 267

Db 587 DRFENDGLIKIREYLVNDIE---PSTFMEFTWCKHDATGTRVLD-KTVSGSISAE- 641

QY 268 VEVEICKIQCEINLTPYIGIPEKRVATMYH-----YLOFATYKL-----NDKNAAPC 317

Db 642 -DVLE---HLDSSQDDVLIDLVKNSVT--HRAALLKLKLVNKLIGSLDHNDRKRLT 695

QY 318 AVSYLTFDQNDKVMQONLVYYOYHRDTWG--LSDEHQPFRPAVOFNVVTTLOKELYDFAK 376

Db 696 AIEF-----RELPCMEKRYTMOYLEQLMGDMSD-----FTHLSFYILCS 736

QY 377 ENIMDDGEVEYVDDL 395

Db 737 INLMENVEESIDKIKDIL 755

RESULT 10

C81297

proline dehydrogenase (EC 1.5.99.8) / 1-pyrroline-5-carboxylate dehydrogenase (EC 1.5

C:Species: Campylobacter jejuni

C&gt;Date: 31-Mar-2000 #sequence,revision 31-Mar-2000 #text,change 15-Sep-2000

C:Accession: C81297

R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chl

C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Ba

Nature 403, 665-668, 2000

A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals

A:Reference number: A81250; MUID:20150912

A:Accession: C81297

A&gt;Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1162 &lt;PAR&gt;

A:Cross-references: GB:AL139078; GB:AL111168; NID:96968723; PIDN:CAB73924.1; PID:9696

A:Experimental source: serotype O2, strain NCTC 11168

C:Genetics:

A:Gene: putA; Cj1503c

C:Keywords: oxidoreductase

Query Match 4.6%; Score 97.5; DB 2; Length 1162;

Best Local Similarity 23.0%; Pred. No. 13; Mismatches 113; Indels 49; Gaps 15;

Matches 62; Conservative 46; Mismatches 113; Indels 49; Gaps 15;



149 KPIQFAVFKANLPAKATAAHTFLKHPDDEMKRNMAYYKSLPGADYIKD----- 201  
Db 217 KFTNLMEERDLELTVAAAMESAKD---IKAGVLQAYIPDSKYIKLPAFSEKER 272  
QY 202 ETSYSLEFTRAVAYNGENMRTSITPDELALPDFEKAFA-----YECLA--ACEGSEIK 254  
Db 273 VLGMKSIRKRFVKGAMMESEETIASQGWELPFTFYKIKDTDSNKNMDFILEGD---- 328  
QY 255 DEKDFYISADHYV-EVLECKIOCEENLTPVIGGYPEKVFATMYH--YLQFAY--YKLN 309  
Db 329 NRYKINIGIASHNIFFELAYVTRISE-----VGA--LDSFTFEMLEGMSIQCSYEELSKM 381  
QY 310 DLKNAAP-C-----AVSYLLFDQNDKVMQONLYYQYH-----RDTWGLSDEHFPQRE 357  
Db 382 DLILYFPVCEAHFNNAIATVLRRLDENISEDMRIFENLKVGDKNMNIQKLELKSLE 441  
QY 358 AVQFNVYTLQKELYDFAKE-NIMDDDEGE 386  
Db 442 GIKTLDTHTKQ--DNKESNIISSEK 469

RESULT 11  
A:71076  
hypothetical protein PH0873 - Pyrococcus horikoshii  
C:Species: Pyrococcus horikoshii  
C:Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 05-Nov-1999  
C:Accession: A71076  
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, S.; Sekit  
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi  
DNA Res. 5, 55-76, 1998  
A:file: Complete sequence and gene organization of the genome of a hyper-thermophilic  
A:Reference number: A71000; MUID:98344137  
A:Accession: A71076  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-788 <RAM>  
A:Cross-references: GB:AP000004; NID:g3236131; PIDN:BAA2967.1; PID:d1030910; PID:g32572  
A:Experimental source: strain OT3  
A:Note: this accession replaces an interim accession for a sequence replaced by GenBank  
C:Genetics:  
A:Gene: PH0873

Query Match 4.5%; Score 96; DB 2; Length 788;  
Best Local Similarity 24.4%; Pred. No. 11;  
Matches 64; Conservative 42; Mismatches 98; Indels 58; Gaps 15;  
QY 153 FAYFKANLPAKATAAHTFLKHPDDEMKRNMAYYKSLPGADYIKDLETKSYESLIR 212  
Db 429 FIVYKAPAEBCRKMSRPYIFLSVRPDMDSKLLGYVNPITGKNKTKLD-----FTWQ 482  
QY 213 AVRAY--NGEN---WRSTIDMELALPDFEKAFAECLACGSEIKDEFYISADHY 267  
Db 483 AIEDYKKNNGNALPYIILIDEMNLAHVEY--FADPLSVLESGDEGFT--RESIKLD 538  
QY 268 VEVLCK--LOCENLTP---VIGGYPEKVFATMYH-----LOFAY--YKLNKNA 315  
Db 539 VDEEKEQGPKEIKLPNLYIGTVNIDE---TTTFESKVLDRAYIEHEVDLENYP 595  
QY 316 PCAVSYLLFDQNDKVMQONLYYQYHRTWGLSDEHFPQREAVQFN-----VTLQKE 370  
Db 596 PREBK--LDEHVVVALRNILIEDLRDG-----KFLNYSKHEINEAVRE 637  
QY 371 LYPAKENIMDDDEGEVVEYVD 392  
Db 638 L-DLSKVRILN---EVLPEYD 654

RESULT 12  
E90139  
conserved hypothetical protein [imported] - Sulfolobus solfataricus  
C:Species: Sulfolobus solfataricus

C:Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 24-May-2001  
C:Accession: E90139  
R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Ameyez, M.J.; Ci  
Jong, I.; Jelifies, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redde  
arrett, R.A.; Ragan, M.A.; Senses, C.W.; Van der Oost, J.  
submitted to GenBank, April 2001  
A:Description: Sulfolobus solfataricus complete genome.  
A:Reference number: A99139  
A:Accession: E90139  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-240 <KUR>  
A:Cross-references: GB:AE006641; NID:g13813132; PIDN:AAK40372.1; GSPDB:GN00155  
C:Genetics:  
A:Gene: SSO0007

Query Match 4.5%; Score 95.5; DB 2; Length 240;  
Best Local Similarity 23.8%; Pred. No. 2.4;  
Matches 54; Conservative 36; Mismatches 88; Indels 49; Gaps 10;  
QY 175 HPD-----DEMKRN---MAYKSLPGADYIKDLETKSYESLIFRAVRAYNGEN----- 221  
Db 36 HPDQWGLSIFLKRNDKMKVITYRS---EDPIYEDPNISY-----LIDAKGEDIVGI 85  
QY 222 ---WRSTIDMELALPDFEKA-FYECLAACEGSEIKDEFKDFYLSIADHYEVLECKI 275  
Db 86 HARKAGKFLTGLSHAPYIRANITYLFRHNSVSRTSKDSNRPFTSYLLEKRT 145  
QY 276 QCEENLTPVIGGYPEKVFATMYHLYQFAYKLNKNAAPCAVSYLLFDQNDKVMQONT 335  
Db 146 LIESNM--PDVAI-----TLDKLDGCSLSLSLIL--YNNKREGPSIL 187  
QY 336 VYQYHNDTWG--LSDHFPQREAVQFNVTTLQKELYDFAKKNIM 380  
Db 188 IGYVYNNRNLSEKTEINERYKLYTDNKRYPFSSTIKYLVGDAEELM 234

RESULT 13  
H65011  
probable transaminase (EC 2.6.1.-) b2379 [similarity] - Escherichia coli (strain K-1  
C:Species: Escherichia coli  
C:Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 17-Mar-2000  
C:Accession: H65011  
R:Blatner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.  
.A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A:file: The complete genome sequence of Escherichia coli K-12.  
A:Reference number: A64720; MUID:97426617  
A:Accession: H65011  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-412 <BLAT>  
A:Cross-references: GB:AE000326; GB:U00096; NID:g1788718; PIDN:AAC75438.1; PID:g1788  
A:Experimental source: strain K-12, substrain MG1655  
C:Superfamily: aspartate transaminase  
C:Keywords: aminotransferase; phosphoprotein; pyridoxal phosphate  
P;244/Binding site: pyridoxal phosphate (Lys) (covalent) #status predicted

Query Match 4.5%; Score 95; DB 2; Length 412;  
Best Local Similarity 19.8%; Pred. No. 5.4;  
Matches 77; Conservative 54; Mismatches 126; Indels 132; Gaps 21;  
QY 15 LCYACALRAGRAQYERS-FRSPFRDELMPLESAYRIAL-----DKTSGEHMAESXYLE 68  
Db 58 LCTYV---ACRPDTHGYSYSGIPIR-----LRRALISRWYODRYVDEIDSESAIVT 104  
QY 69 ISLR---LHRLDSEAFCHRNCSAQPQEPAGLAGLYPELRLFGGLRRAHCKRKQK 125  
Db 105 IGSKEGIALMLATLD---IGDYLVPNP-----STP-IHTYGAIVAGAYV-----RS 148  
QY 126 LP-----AFRQSQPS-REVLADFORREPYKFLQFAVFKANLPAKATAAHTF 171

Db 149 VPLVEGVDFNLELRAIRRESYPRKPMILIGFSPNPTAOCVELEFFE-----KVALARXY 203  
QY 172 LKHHDDMMKRNMA-----YKSLPGAEDYIKDLETSYSLSFIRAVRAYNG 219  
Db 204 -----DVLVYHDLVIAITVDGKAPSIMOYPGARDVAVEFTLS-----KSTNM 248  
QY 220 ENMRTSTIDMELAPDF---FKAFYE-----CLACGSGRE-IKDFKDFYLSIAD 265  
Db 249 AGWRIGFVWGNKTLVSALARIKSYHDGTFPLQVAATAALEGGDQCVRDIAEQYKRRRD 308  
QY 266 HVEVLECKIQCEENLTIVIGYPRYKFAVATYAH-----IQFAYTKLNDLK- 312  
Db 309 VLVKGLH-----EAGMVMEMPRKASYVMAKIPEPYAANGSLEFAKKILNEAKV 356  
QY 313 -----NAAPCAVSYLFDONDKVMQ 332  
Db 357 CVSPGICRGDYGDTHTVRFALLENDRIRQ 385

## RESULT 14

S64361  
hypothetical protein YGR066c - yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein G4545  
C:Species: Saccharomyces cerevisiae  
C>Date: 17-May-1996 #sequence\_revision 17-May-1996 #text\_change 29-Oct-1999  
C:Accession: S64361  
R:Medler, H.; Scharfe, M.; Medler, E.; Wambutt, R.  
submitted to the Protein Sequence Database, May 1996  
A:Reference number: S64356  
A:Accession: S64361  
A:Molecule type: DNA  
A:Residues: 1-292 <MED>  
A:Cross-references: EMBL:Z72851; NID:q1323087; PID:e243420; PID:q1323088; GSPDB:GN00007;  
A:Experimental source: strain S288C  
C:Genetics:  
A:Gene: MIPS:YGR066c  
A:Map position: 7R

Query Match 4.4%; Score 94.5; DB 2; Length 292;  
Best Local Similarity 23.2%; Pred. No. 3.8;  
Matches 63; Conservative 29; Mismatches 87; Indels 93; Gaps 12;  
QY 179 EMMKRNMAVYKSLPGAEDYIKDLETSYSLSFIRAVRAYNGENMRTSTIDMELAPDFK 238  
Db 14 ERAKRN-----DNIEIYTSAYVSL-----PSTDFR--LPHFKA 46  
QY 239 AFYECGLACGSGREI-KDFKDFYLSIADH-----YEVLECKIQCE 278  
Db 47 KEEDCDVYEECTNLVGNKAKYTRSLGRHDLFRGLRFGSGSSKYYTYTV-EYKIDTV 105  
QY 279 -----ENLTPVIGYPRYKFAVATYAHYLOFAVYKLNLDKNAAPC 317  
Db 106 NLPLXKDSRSLDPHTGTFITKNTLPVLDKY-VTLFEGYVINYQFPL-----C 153  
QY 318 AVSYLFDONDKVMQON-----LVYYQYHRDTWGLSDEHF-QPRPAVOFFNVTTLQK 369  
Db 154 SLHMPAEETLDPYMAQRESDCSHMKRFGHFGSDNMSLTERNFQYNHESAPEFMN---QR 209  
QY 370 ELVDFAKENIMDDDEGEVVEVYVDLLELETS 401  
Db 210 YIYLKMKERFLDDEEDENIMLDNHHLEGAS 241

## RESULT 15

B81279

probable fibronectin/fibrinogen-binding protein Cj1349c [Imported] - Campylobacter jejuni  
C:Species: Campylobacter jejuni  
C>Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 31-Mar-2000  
C:Accession: B81279  
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Kelley, J.M.; Churcher, C.; Basham, D.; Chilling  
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barr

Nature 403, 665-668, 2000  
A>Title: The genome sequence of the food-borne pathogen Campylobacter jejuni, reveals  
A:Reference number: A81250; MUID:20150912  
A:Accession: B81279  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-435 <PAR>  
A:Cross-references: GB:AL139078; GB:AL111168; NID:q6968723; PID:CA873776.1; PID:q696  
A:Experimental source: serotype O2, strain NCTC 11168  
C:Genetics:  
A:Gene: Cj1349c

Query Match 4.4%; Score 94; DB 2; Length 435;  
Best Local Similarity 20.1%; Pred. No. 7;  
Matches 59; Conservative 54; Mismatches 87; Indels 94; Gaps 14;  
QY 135 SREVLADFORREPYKFLQFAIFKAN---NLKRAIAAHT---FLKH---PDDMKKRNMA 186  
Db 14 SRFKKIDFIKRINDNILELSPDRERFIPLDTRAMSAITAKFNKYNAPDFMLK--- 70  
QY 187 YKSLPGAEDYIKDLE-----TKSYESLFIIRAVRAYNGENMRTSTID- 228  
Db 71 YFSNA-----FLKEVYKLEGNRIILCFSVKANAKASYESKTYFEPTGANTVILTDKDL 125  
QY 229 -----MELALPDFFAFYECGLACGSGREIKDFYLSIADHYEVLECKIQ 277  
Db 126 IIEALRHIDKSYRVKRPVILLESLSKPKYKMDENFEIEIKDFADYF---SRKFTSYESKIKQ 182  
QY 278 EENLTPVIGYPRYKFAVATYAHYLOFAVYKLNLDKNAAPCAVSYLFDONDKVMQONLY 337  
Db 183 IKNLKLA---QVDK-----KIONLOEL-----FSLDEENSLKLKALEY 218  
QY 338 YQYHNDTWGLSDEHQRPRPA-VQFENVYTLQKELYDFAKENIMDDDEGEVVEY 390  
Db 219 -----RKRADVLEFANLSYLK---DIREFEKLDDEGKELEF 251

Search completed: June 8, 2002, 02:09:53  
Job time: 6933 sec





SEA ID NO: 1  
Database: N-Geneseg-032802  
AC NO: AAX00801File Copy  
Page 10GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 7, 2002, 21:01:30 ; Search time 562.42 Seconds  
(Without alignments)  
11817.104 Million cell updates/secTitle: US-09-729-674-1  
Perfect score: 3871  
Sequence: 1 ttccctctccctccctt.....ataaaaaaaaaaaaaa 3871Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapect 1.0Searched: 1736436 seqs, 858457221 residues  
Total number of hits satisfying chosen parameters: 3472872Minimum DB seq length: 0  
Maximum DB seq length: 2000000000Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summariesDatabase :  
1: N-Geneseg\_032802.\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
25: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3870.2	100.0	3871	20	AAX60801 Human secreted pro
2	3870.2	100.0	3871	22	AAS59207 Human cDNA encodin
3	3870.2	100.0	3871	24	ABA90876 Human polynucleoti
4	2051.2	53.0	2549	21	AAFI6225 Human prostate can
5	1955.4	50.5	1961	22	AAH14525 Human cDNA sequenc
6	1474.4	38.1	1586	21	AAH31138 Human cDNA clone (
7	695.6	18.0	818	22	AAH06808 Human colon cancer
8	466.4	12.0	618	21	AAH80131 Human colon cancer
9	406.4	10.5	408	21	AAH30585 Human colon cancer

10	371.6	9.6	401	22	AAH50786 Human tumour assoc
11	369.4	9.5	400	16	AAH21281 Human gene signatu
12	355.6	9.2	407	22	AAH50808 Human tumour assoc
13	340	8.8	351	21	AAH11840 Human cDNA clone (
14	337.2	8.7	373	21	AAH13677 Human secreted exp
15	258.6	6.7	279	23	AAH58007 cDNA #683 encoding
16	239.2	6.2	6670	22	AAH02846 Human reproductive
17	239.2	6.2	7713	22	AAH02846 Human reproductive
18	238.4	6.2	8658	22	AAH69231 Human immune/haema
19	238	6.1	295	22	AAH82995 Human immune/haema
20	235.4	6.1	32351	21	AAH21307 Human low adenosin
21	235.4	6.1	32351	21	AAH35185 Human adenosine re
22	235.4	6.1	40298	21	AAH21311 Human adenosine re
23	235.4	6.1	40298	21	AAH35189 Human nervous syst
24	234.6	6.1	4275	22	AAH19408 Human immune/haema
25	234.6	6.1	4883	22	AAH79410 Human immune/haema
26	234.6	6.1	16997	22	AAH80033 Human immune/haema
27	234.6	6.1	16997	22	AAH86146 Human immune/haema
28	233.8	6.0	9620	22	AAH06207 Human reproductive
29	233.4	6.0	6138	22	AAH57456 Human liver cell s
30	233.2	6.0	763	22	AAH06411 Human cDNA clone (
31	232.6	6.0	2263	22	AAH14438 Human CDNA sequenc
32	232.6	6.0	5979	22	AAH78985 Human immune/haema
33	232.6	6.0	5979	22	AAH80951 Human immune/haema
34	232.2	6.0	532	22	AAH71264 Human cervical can
35	232.2	6.0	567	22	AAH69839 Human polynucleoti
36	232	6.0	3775	22	AAH61061 Human immune/haema
37	231.4	6.0	1449	22	AAH86057 Human immune/haema
38	231.4	6.0	1450	22	AAH86062 Human nervous syst
39	230.8	6.0	2847	22	AAH18224 Human immune/haema
40	230.8	6.0	2847	22	AAH82073 Human immune/haema
41	230.6	6.0	2108	22	AAH76677 Human immune/haema
42	230.6	6.0	2108	22	AAH76677 Human immune/haema
43	230.2	5.9	646	22	AAH79815 Human digestive sy
44	230.2	5.9	11446	22	AAH90464 Human nervous syst
45	230	5.9	13559	22	AAH15144 Human

## ALIGNMENTS

RESULT 1	
AAH60801	
ID	AAH60801 standard; DNA: 3871 BP.
XX	
XX	AAH60801:
XX	
XX	09-AUG-1999 (first entry)
XX	
XX	Human secreted protein encoding DNA (clone bd306-7).
XX	
XX	Secreted protein; kidney; lung; brain; blood; testis; bone marrow;
XX	nutritional activity; cytokine; cell proliferation; immune stimulation;
XX	hematopoiesis regulation; tissue growth; thrombolytic; gene therapy;
XX	anti-inflammatory; tumour invasion; ss.
OS	
XX	Homo sapiens.
XX	
XX	WO926961-A1.
XX	
XX	03-JUN-1999.
XX	
XX	24-NOV-1998; 98MO-US25149.
XX	
XX	23-NOV-1998; 98US-0197886.
XX	
XX	26-NOV-1997; 97US-0066804.
XX	
XX	(GEMT) GENETICS INST INC.
XX	
XX	Agostino MJ, Clark HF, Collins-Racie LA, Evans C;
XX	Fechel K, Jacobs K, Lavallee ER, McCoy JM, Metberg D;
XX	Steininger RJ, Treacy M, Wong GC;



[illegible]

QY	601	tgtatgaagaggaacaatgcatattataaagagccttgctgtgttcgcggagacataataaag	660
Db	601	tgtatgaagaggaacaatgcatattataaagagccttgctgtgttcgcggagacataataaag	660
QY	661	acctcgaaaccaaatctcaataagaaagcctgtgtctaccgaagcagtcgagatacaaacggtg	720
Db	661	acctcgaaaccaaatctcaataagaaagcctgtgtctaccgaagcagtcgagatacaaacggtg	720
QY	721	agaacvtgagaaacataccatacagaagatvgagcttgccttcccgactctctcaaacgct	780
Db	721	agaacvtgagaaacataccatacagaagatvgagcttgccttcccgactctctcaaacgct	780
QY	781	tttacgagvtgtcttcgcagccttcgagaggtgtccagagagatcaagatctcaagattctc	840
Db	781	tttacgagvtgtcttcgcagccttcgagaggtgtccagagagatcaagatctcaagattctc	840
QY	841	accttccatagcagatcatatgataagaaatcttgaaatgcaaaataacagttgtgaagaa	900
Db	841	accttccatagcagatcatatgataagaaatcttgaaatgcaaaataacagttgtgaagaa	900
QY	901	acctcaacccccagttbatagagagcgtatccggttgtagaaattgtgtgcatacctatgatac	960
Db	901	acctcaacccccagttbatagagagcgtatccggttgtagaaattgtgtgcatacctatgatac	960
QY	961	acttgcaagtttgcccttatataaagtgtgaacgacctgaagaaatgacgccccctgttgacga	1020
Db	961	acttgcaagtttgcccttatataaagtgtgaacgacctgaagaaatgacgccccctgttgacga	1020
QY	1021	gctatcgcgtctcttgatcaacaatgtaaaaggtcatgacagaaagaaacctggtgtatccagt	1080
Db	1021	gctatcgcgtctcttgatcaacaatgtaaaaggtcatgacagaaagaaacctggtgtatccagt	1080
QY	1081	accacaagagacacttgggagcctctcgatgtagacacttccagccccagaacctgtgaagcattc	1140
Db	1081	accacaagagacacttgggagcctctcgatgtagacacttccagccccagaacctgtgaagcattc	1140
QY	1141	agttctttaatgtgacaacaacacctccaagaagagatgtatgactttgttaagagaaatatata	1200
Db	1141	agttctttaatgtgacaacaacacctccaagaagagatgtatgactttgttaagagaaatatata	1200
QY	1201	tggatgtatgatatgaagaggaagctgtctgaaatatgtggatgaacctctggaacctggaagaa	1260
Db	1201	tggatgtatgatatgaagaggaagctgtctgaaatatgtggatgaacctctggaacctggaagaa	1260
QY	1261	ccagactagcccaacagcaaccaaaagaaactctcctctgtgaggttccaagaaacacagaattcct	1320
Db	1261	ccagactagcccaacagcaaccaaaagaaactctcctctgtgaggttccaagaaacacagaattcct	1320
QY	1321	tgtctcttttcccaacagcccaagcctgtgtgataacctagaagccttctcttaactctccaaa	1380
Db	1321	tgtctcttttcccaacagcccaagcctgtgtgataacctagaagccttctcttaactctccaaa	1380
QY	1381	gtgaaagaggaagcccccggtctctctcaacgtatcatcaagggtgtgaagcgtgccttccct	1440
Db	1381	gtgaaagaggaagcccccggtctctctcaacgtatcatcaagggtgtgaagcgtgccttccct	1440
QY	1441	atcttcaacacctgcaacctcatgtgtcaacaacctattctcaaccttttttttgagatgta	1500
Db	1441	atcttcaacacctgcaacctcatgtgtcaacaacctattctcaaccttttttttgagatgta	1500
QY	1501	gtctcgcgtctcttcgcccagcctgagatgcaatgcaatgcaagtttccagctcaatgcacacctcc	1560
Db	1501	gtctcgcgtctcttcgcccagcctgagatgcaatgcaatgcaagtttccagctcaatgcacacctcc	1560
QY	1561	gctcctctgggttcaagaatctctgcgtcatcaagcccccggaatcacctgagatgaatgaagca	1620
Db	1561	gctcctctgggttcaagaatctctgcgtcatcaagcccccggaatcacctgagatgaatgaagca	1620
QY	1621	tgtgtccacacacgcccgcgtcaattttgtatttttagtataagaaaggggttttgcaagtttgg	1680
Db	1621	tgtgtccacacacgcccgcgtcaattttgtatttttagtataagaaaggggttttgcaagtttgg	1680
QY	1681	ccagagctggtctcgaactcttgacttcaatgataatgatacattgctcctgtgcctccacagatgac	1740

**THIS PAGE BLANK (0.3PT0)**



Db 1681 ccaagcgctgcgaactcttgacttcaagatgatacctcgtccttgcctcccaaatgctc 1740  
QY 1741 tgggattacagcgctgagcacaatgcccggcctcttctcaaccttacaacctgtctct 1800  
Db 1741 tgggattacagcgctgagcacaatgcccggcctcttctcaaccttacaacctgtctct 1800  
QY 1801 taccctcaactcgttttcaacaccttcaacctcgtctctctcaatgttcaacctgtctc 1860  
Db 1801 taccctcaactcgttttcaacaccttcaacctcgtctctctcaatgttcaacctgtctc 1860  
QY 1861 cccatgttcaatgctgccttcttcaaccttcttggttgaaggcagctcctctgtgttg 1920  
Db 1861 cccatgttcaatgctgccttcttcaaccttcttggttgaaggcagctcctctgtgttg 1920  
QY 1921 ttttttcttcccaagaaatcagatattttttaaataagaaataacattcctagaa 1980  
Db 1921 ttttttcttcccaagaaatcagatattttttaaataagaaataacattcctagaa 1980  
QY 1981 gatgataatgttgaataacacctcttggcttattgtcttccagattttagctccttc 2040  
Db 1981 gatgataatgttgaataacacctcttggcttattgtcttccagattttagctccttc 2040  
QY 2041 tcccattccgggaagaatgttgaagacataagcttaattctccagcctcacaatgctc 2100  
Db 2041 tcccattccgggaagaatgttgaagacataagcttaattctccagcctcacaatgctc 2100  
QY 2101 ttcaacttgcctgactgttcaaccttcaacctcagcctgaataaagaatgttgaagag 2160  
Db 2101 ttcaacttgcctgactgttcaaccttcaacctcagcctgaataaagaatgttgaagag 2160  
QY 2161 tgtagaagtcgaagaatgttgccttgcctcccaacttgcattcccaaatlaacaagcttg 2220  
Db 2161 tgtagaagtcgaagaatgttgccttgcctcccaacttgcattcccaaatlaacaagcttg 2220  
QY 2221 ccgattcccaatttgaagacaatgttgaataagcttcgagcttgaagaaaggaagc 2280  
Db 2221 ccgattcccaatttgaagacaatgttgaataagcttcgagcttgaagaaaggaagc 2280  
QY 2281 cagagctgtcctgaattcattcattcttcaagtaattattatgttgaattacattgtgtg 2340  
Db 2281 cagagctgtcctgaattcattcattcttcaagtaattattatgttgaattacattgtgtg 2340  
QY 2341 taagcaatgaacctggaactagaagatactcacagaataacagaggaagttccctgtgt 2400  
Db 2341 taagcaatgaacctggaactagaagatactcacagaataacagaggaagttccctgtgt 2400  
QY 2401 catgagcttactaattcttaacaggaagaaagataagccaatacataggaataatataac 2460  
Db 2401 catgagcttactaattcttaacaggaagaaagataagccaatacataggaataatataac 2460  
QY 2461 aaggtatcatgtatgataatctgttggagaataataagcagaggaaggttaagaa 2520  
Db 2461 aaggtatcatgtatgataatctgttggagaataataagcagaggaaggttaagaa 2520  
QY 2521 tccctgagagaagaagctgcaagtttaaatgagcctcactcaggaatgtgaagcaga 2580  
Db 2521 tccctgagagaagaagctgcaagtttaaatgagcctcactcaggaatgtgaagcaga 2580  
QY 2581 gacgtttaggagaagtgatctcctgacagaagcattccagcagaggaagaagatgtgactg 2640  
Db 2581 gacgtttaggagaagtgatctcctgacagaagcattccagcagaggaagaagatgtgactg 2640  
QY 2641 ccccaagaaggaactgtcttactgctgaggaaggaagcaggaagcacaagatgctg 2700  
Db 2641 ccccaagaaggaactgtcttactgctgaggaaggaagcaggaagcacaagatgctg 2700  
QY 2701 tgggacagaggttagaataaggaagagagcgctgaggaagacagagtgtagagccttg 2760  
Db 2701 tgggacagaggttagaataaggaagagagcgctgaggaagacagagtgtagagccttg 2760  
QY 2761 cttctgctaagtgagatggaacacactgagaggttgaacagagagtgcttgatg 2820

Db 2761 cttctgctaagtgagatggaacacactgagaggttgaacagagagtgcttgatg 2820  
QY 2821 ttatattttgcaaggggctcattcagctgcaataatttggaataaactttagtggaagagc 2880  
Db 2821 ttatattttgcaaggggctcattcagctgcaataatttggaataaactttagtggaagagc 2880  
QY 2881 agagaagaagaggaagacactgtttagaagaagctactgcaaggttccagagcttgagcctg 2940  
Db 2881 agagaagaagaggaagacactgtttagaagaagctactgcaaggttccagagcttgagcctg 2940  
QY 2941 cacagcaacagcgctgttcaaatatcagatattattttgaagaagcacaataagatttc 3000  
Db 2941 cacagcaacagcgctgttcaaatatcagatattattttgaagaagcacaataagatttc 3000  
QY 3001 tgaagatttgaatgtgagtgtaagaagaagaagattttagatgacattaaagttttg 3060  
Db 3001 tgaagatttgaatgtgagtgtaagaagaagaagattttagatgacattaaagttttg 3060  
QY 3061 gctcgaatagcaggaagaagatgagttacagatcagtaataaggaagagatgagctggt 3120  
Db 3061 gctcgaatagcaggaagaagatgagttacagatcagtaataaggaagagatgagctggt 3120  
QY 3121 aagtaaggaattgtgtgcaagaagcagctgtctgtgttggaagagaggttctgctgca 3180  
Db 3121 aagtaaggaattgtgtgcaagaagcagctgtctgtgttggaagagaggttctgctgca 3180  
QY 3181 aatcaagaatgagatcctcctcagatcagctgctcagcagagcctcgaagaagatcagaa 3240  
Db 3181 aatcaagaatgagatcctcctcagatcagctgctcagcagagcctcgaagaagatcagaa 3240  
QY 3241 tgcacttgggtttaaattgttgggggtgctcctcagaaggaagacactgtgaagccttatacgc 3300  
Db 3241 tgcacttgggtttaaattgttgggggtgctcctcagaaggaagacactgtgaagccttatacgc 3300  
QY 3301 attatttgcctgtgagaagctcctcctgagagtggtgtgtaactttgaagcagaatgacttca 3360  
Db 3301 attatttgcctgtgagaagctcctcctgagagtggtgtgtaactttgaagcagaatgacttca 3360  
QY 3361 gttgaaggaagatcctcctcgaagaagagcgttagacatctgcaagcctacatctgtgtgtag 3420  
Db 3361 gttgaaggaagatcctcctcgaagaagagcgttagacatctgcaagcctacatctgtgtgtag 3420  
QY 3421 tctgttgggggttgggggtcctgagcactgtgctgtggaagagctcctgagcagcaca 3480  
Db 3421 tctgttgggggttgggggtcctgagcactgtgctgtggaagagctcctgagcagcaca 3480  
QY 3481 ggcgccctactgaaacatcagcatgtcagttgacatttaagccatgagctggaagggc 3540  
Db 3481 ggcgccctactgaaacatcagcatgtcagttgacatttaagccatgagctggaagggc 3540  
QY 3541 cactgagattgtcctcgtgaattactcaggaagaacaagaagaagacatgagatgagcc 3600  
Db 3541 cactgagattgtcctcgtgaattactcaggaagaacaagaagaagacatgagatgagcc 3600  
QY 3601 ttgggctcctcctggaagaatggaatcagcgaaggaagcctgaagaaggtttaaacttaaggtc 3660  
Db 3601 ttgggctcctcctggaagaatggaatcagcgaaggaagcctgaagaaggtttaaacttaaggtc 3660  
QY 3661 agagaanaacaaagagatgtgtgtctgtggaagctgaggtttcttatcaacactcattc 3720  
Db 3661 agagaanaacaaagagatgtgtgtctgtggaagctgaggtttcttatcaacactcattc 3720  
QY 3721 ccttctccaaataaagcactgtgtgagttggtggtccctccaggttgaagcagaagagaga 3780  
Db 3721 ccttctccaaataaagcactgtgtgagttggtggtccctccaggttgaagcagaagagaga 3780  
QY 3781 aaggaacagcgcttgggaagaagaagcatttccctgcaataagccttgggaagaagaaga 3840  
Db 3781 aaggaacagcgcttgggaagaagaagcatttccctgcaataagccttgggaagaagaaga 3840  
QY 3841 tagagtggttaataataaaaaaa 3871  
Db 3841 tagagtggttaataataaaaaaa 3871

THIS PAGE BLANK (EXPTO)



THIS PAGE BLANK (U.S.P.T.O.)

**SQ Sequence 401 AA;**

QY	1	MEPBRGAALIALALCYACALRGRPAQERYSPFSFPRDLMPESRYRRAADKYSEHW	60
Db	1	mepirgaaliallcvacalragraqerysfstfcdelmpesrynaldkyseshw	60
QY	61	AESGYEIELRLRRLRDSEAFCHNRCSNAOPPEPAGIASYPELIRGGLRRACIK	120
Db	61	aesgyeielrlhrrlrdseafchmcsapqepagiasypelirgllrracik	120

[illegible]

RESULT	2
AAU38989	
ID	AAU38989 standard; Protein; 401 AA
XY	

DT 16-JAN-2002 (first entry)

Human secreted protein bd306\_7.

Human: secreted protein; antiinflammatory; immunosuppressive; neurotropic; neuroprotective; antitachycardic; antimicrobial; vulnerrary; cytotstatic; antidiabetic; vitruclide; antifertility; anticonvulsant; vasotropic; antiparkinsonian; immunostimulant; dermatological; antihemmatic; antitumor; anticancer; osteopathic; tranquilliser; cerebroprotective; cytokine; cell proliferation; cell differentiation; immune deficiency; severe combined immunodeficiency; SCID; tumour; autoimmune disorder; multiple sclerosis; rheumatoid arthritis; graft-versus-host disease; myeloid deficiency; wound healing; ulcer; periodontal disease; osteoporosis; osteoarthritis; Alzheimer's disease; Parkinson's disease; Huntington's disease; infection; cardiac disease; stroke; sepsis; inflammatory bowel disease; contraceptive; immunogen; food supplement; vaccine.

Homo sapiens.

WO200175068-A2.

11-OCT-2001.

22-MAR-2001; 2001WO-US09369.

30-MAR-2000; 2000US-0539330.

04-DEC-2000; 2000US-0729674.

(GEMX ) GENETICS INST INC.

Jacobs K, McCoy JM, Lavallie E, Collins-racie LA, Evans C; Treacy M, Agostino MJ, Steininger RJ, Spaulding V, Wong GG; Clary H, Fechtel K, Werberg D; WPI; 2001-639363/73.

N-PSDB; AAS59207.

Secreted human proteins, useful as vaccine for treating various diseases such as autoimmune disorders (e.g. multiple sclerosis), and nervous system disorders (e.g. stroke) -

Claim 2; Page 457-459; 619pp; English.

The invention relates to novel human secreted proteins, the nucleic acids encoding them. The protein may exhibit cytokine, cell proliferation or cell differentiation activity or may induce production of other cytokines in certain cell populations and may exhibit immune stimulating, or immune suppressing activity, which is useful for the treatment of - various immune deficiencies and disorders e.g. severe combined immunodeficiency (SCID), autoimmune disorders e.g. multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation. The proteins are also useful in the treatment of diseases and disorders including tissue, skin and organ transplantation and in graft-versus-host diseases (GVHD). In the induction of tumour immunity, myeloid or lymphoid cell deficiencies, wound healing and tissue repair, in the treatment of burns, incisions and ulcers, as well as in treatment of periodontal disease, osteoporosis or osteoarthritis, mediated by inflammatory processes, diseases of the peripheral nervous system, Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome, infections, infarction of cardiac and central nervous system vessel e.g. stroke, sepsis, inflammatory bowel disease, ulcers, bone regeneration. The protein, having activin- or inhibin-related activities is useful as a contraceptive based on the ability of inhibins to decrease fertility in female mammals and decrease spermatogenesis in male mammals. The proteins and nucleic acids are also useful as food supplements. The present sequence represents a secreted protein of the invention.

THIS PAGE BLANK (csp70)

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 7, 2002, 20:54:15 ; Search time 3903.73 Seconds  
(without alignments)  
13383.793 Million cell updates/sec

Title: US-09-729-674-1  
Perfect score: 3871  
Sequence: 1 ttccctctccctccctt.....aataaaaaaaaaaaaaaa 3871

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hcc:\*  
9: gb\_estl:\*  
10: gb\_estl2:\*  
11: gb\_hcc:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	996.4	25.7	1079	9	AL569764
2	993.8	25.6	1029	9	AL545483
3	964.4	24.9	980	9	AL517454
4	954.4	24.7	1031	9	AL571267
5	936.2	24.2	964	9	AL546910
6	931.6	24.1	971	9	AL556976
7	924.4	23.9	1652	11	AK017797
8	918	23.7	971	9	AL572160
9	917.8	23.7	1005	9	AL514488
c 10	917.4	23.7	925	9	AL574681
11	896.6	23.2	1053	10	BM471592
12	889.6	23.0	916	9	AL541167
13	860	22.2	909	9	AL517455
14	858.8	22.2	973	9	AL543090
15	849.8	22.0	956	9	AL554750
16	816	21.1	916	9	AL550966
c 17	814.8	21.0	830	9	AL579089

c 18	804.8	20.8	863	9	AL580676
19	788	20.4	961	10	BI255097
20	777	20.1	827	10	BI760650
21	772.6	20.0	817	9	AL558186
22	771.4	19.9	774	9	AL598447
23	770.6	19.9	854	10	BI860855
24	765.2	19.8	846	10	BI253401
25	760.8	19.7	813	9	AL558966
26	759.6	19.6	933	10	BI858585
27	742.8	19.2	797	9	AU141715
28	741.6	19.2	1034	10	BM459357
29	737	19.0	846	10	BI260161
30	736.6	19.0	1088	10	BF025893
31	734.4	18.9	969	10	BF025893
32	731.8	18.9	1430	11	AK007624
33	731	18.9	782	10	BE877856
c 34	720.6	18.6	723	9	AL568770
35	715	18.5	846	10	BI862238
36	713.8	18.4	902	10	BF793387
37	710.6	18.4	934	10	BM453316
38	706	18.2	759	10	BI856439
39	705.2	18.2	707	10	BE747825
40	697.8	18.0	713	10	BE746070
41	695.8	18.0	760	10	BE825941
42	695.6	18.0	769	10	BE828789
43	695.6	18.0	818	9	AU132637
44	693	17.9	702	10	BI256839
45	684	17.7	762	10	BE744751

## ALIGNMENTS

RESULT 1  
AL569764/c 1079 bp mRNA linear EST 16-FEB-2001  
LOCUS AL569764 LTI\_NFL006\_PL2 Homo sapiens cDNA clone CSDDI002Yp17 3  
DEFINITION prime, mRNA sequence.

ACCESSION AL569764  
VERSION AL569764.1 GI:12925427  
KEYWORDS EST.

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE  
1 (Bases 1 to 1079)  
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
Full-length cDNA libraries and normalization

JOURNAL  
Unpublished (2001)  
COMMENT  
Contact: Genoscope

Genoscope - Centre National de Sequencage  
BP 191 91006 Evry cedex - France  
Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr.

## FEATURES

## source

1. 1079  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="CSDDI002Yp17"  
/clone\_1ib="LTI\_NFL006\_PL2"  
/tissue\_type="Placenta"  
/note="Vector: pCMVSPORT 6; Site 1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by life technologies. Contact : Feng Liang Life Technologies, Rockville, Maryland 20850, USA Fax : (1) 301 610 8371  
Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com"

BASE COUNT 317 a 226 c 285 g 246 t 5 others  
ORIGIN

Query Match 25.7%; Score 996.4; DB 9; Length 1079;  
 Best Local Similarity 97.8%; Pred. No. 3.4e-142;  
 Matches 1026; Conservative 4; Mismatches 16; Indels 3; Gaps 2;

```

QY 901 accacacccagttatagaagcgtatccgtttgagaattgtgtctaccatgtacat 960
    ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1046 ACCCTACCCCGATATARK-AGCTATCCAGTGAAGAAAT-TGTGGCTACCATGATCATTT
QY 961 acttcagattgtccatataatgtgaagacccctgaagaatgcacccctgtgcagca 1020
    ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 989 AMTTGAGATTGGCTATATATAGTTGAAGACACCTGTWAAATGACACCCCTGTGACGCA 930
QY 1021 gctatctgctctttgatcagaataagcaagctcatagcagcaaacctgtgtattaccag 1080
    ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 929 GCTATCTGCTCTTTGATGATGATGACAGATGATGACACACACCTGTGTATTAACAGT 870
QY 1081 acccaagagacactttgggctctcgtatgagacacttccagccagacactgaagactc 1140
    ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 869 ACCACAGGACACTTGGGGCTCTCGGATGAGACACTTCCACCCACAGCCTGAAGCAGTTTC 810
QY 1141 agttcttaattgtaccacactccagaagagctgtatgactttgtlaagaaataataa 1200
    ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 809 AGTTCTTAAATGTACCACTCCAGAAAGAGCTGTATGACTTTGCTTAAGAAAATATATA 750
QY 1201 tggatgatgatgagaggaagtgtgtgaataatgtgatgactcttggaaacttgaagaga 1260
    ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 749 TGGATGATGATGAGGAGGAAATGTGAAATGTGATGACCTCTTGAACTGAGAGAGA 690
QY 1261 ccaactgagccacagcaacaaagaagacttctctgtgcttgcaggaagaacagattctt 1320
    ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 689 CCACCTAGCCACACAGCAACAAAGACTTCTCTTGGCTGTCAAGAAACACAGATTTCTT 630
QY 1321 tgtctcttcccaagacccagagctgtgtatcactcagagctcttcttactctccaa 1380
    ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 629 TGTCTTTTCCCAACAGGCCAGGCTGTGATACCTAAGAGCCTCTTTACTCTCCAAA 570
QY 1381 gtgaagaaggagcccccgtctcttaactgaatgtcatcaggggtgagcgtgcttccct 1440
    ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 569 GTGAAAGGGAAGCCCGCTCTCTTACTGATGATGATGATGATGATGATGATGATGATG 510
QY 1441 atcttcaacactgacactcaatgttcaacacttcttcaacttcttcttcttcttctt 1500
    ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 509 ATCTTACACACTGCGACCTATGTTCAACACTTATCTTCTTCTTCTTCTTCTTCTTCT 450
QY 1501 gtctgctctcttgcagagcgtgagtgcaatgagcagcttctcagactcaactgcaactcc 1560
    ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 449 GTCTGCTCTCTTGTCCCAAGCTGAGTGAATGAGTGAATGAGTGAATGAGTGAATGAG 390
QY 1561 ggccttgggttcaagaatctgtgtcatcagcctccagagtaacctggagtaacagga 1620
    ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 389 GCGCTTGGGTTCAAGCAATTTGCTGCTGATGATGATGATGATGATGATGATGATGATG 330
QY 1621 tgtgccaacagccggcctaatttgtatctttagtagagagcgggtttgcaatgttgg 1680
    ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 329 TGTGCCACACAGCCCGGCTAATTTTGTATTTTGTATGATGATGATGATGATGATGATG 270
QY 1681 ccaagcgtgtctcgaaactcttgacttgaatgaatgaatgaatgaatgaatgaatgaat 1740
    ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 269 CCAAGGCTGCTGCTGAACTTGAATGATGATGATGATGATGATGATGATGATGATGATG 210
QY 1741 tgggattaaaggcgtgagccacatgcgcgcgtcttcttcaacttcaacttcaacttct 1800
    ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 209 TGGGATTTACAGGGGTGAGCCACCATGCGCGCTCTTCTTCTTCTTCTTCTTCTTCTT 150
QY 1801 tatcttcaacttgttttcaacttcaacttcaacttcaacttcaacttcaacttcaacttct 1860
    ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 149 TATCTTCACTATGTTTTCACACTTCTATCTGCTTCTTCTTCTTCTTCTTCTTCTTCT 90
QY 1861 cccatgttcaatagcgtcttcttcaacttcttgaaggcagctctctctgagtg 1920
    ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 89 CCATATTTATACCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTG 30

```

QY 1921 ttttttcttcccaagaatcagat 1949  
 Db 29 TTTTCTTATTTTNAATAAATAATATAT 1

## RESULT 2

AL545483 1029 bp mRNA linear EST 16-FEB-2001  
 LOCUS AL545483 LTI\_NFL006.PL2 Homo sapiens cDNA clone CS01015YK22 5  
 DEFINITION prime, mRNA sequence.  
 ACCESSION AL545483  
 VERSION AL545483.1 GI:12877964  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 1029)  
 AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 Evry cedex - France  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES  
 source  
 1..1029  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="CS01015YK22"  
 /tissue\_type="placenta"  
 /note="Vector: pCMVSPORT 6; site\_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 218 a 320 c 285 g 204 t 2 others  
 ORIGIN

Query Match 25.6%; Score 992.8; DB 9; Length 1029;  
 Best Local Similarity 99.2%; Pred. No. 1.2e-141;  
 Matches 1024; Conservative 4; Mismatches 1; Indels 3; Gaps 3;

```

QY 32 gtccctctctctctctcttccgagcgagtgagcggcgagcgagcgagcgagcgagcgagc 91
    ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 GTCCCTTCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 60
QY 92 gctgtcagcgtctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 151
    ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 GCTGCTAGCGCTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 118
QY 152 cttaagcttccgagcgtcttcccaagcgagcgtatgcttccgctgagtcgagcgtcctacgga 211
    ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 119 CTACAGCTTCCGACGCTTCCACGAGGAGACGCTGATCCCTCTGAGTGGCTTACCGGCA 178
QY 212 cgcgctgacaagtagacgagcgagcactgagcgagagcctkagctacactgagatcag 271
    ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 179 CGCGCTGACAGTAGACGCGGACGACTGAGCCGGAAGCCTGGGCTTACTGGAGATCAG 238
QY 272 ccttgagcgtgacacgcttctgtcgcgacagcagagccttctgccaacgcaactgacgagc 331
    ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 239 CTTGGCGGCTGACACGCGTGTGCGGCGACAGC-AGGCTTCTGACACGCAACTGACAGCC 297
QY 332 cgcgcgcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 391
    ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 298 CGGCGCGACGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 357

```



[illegible]

RESULT	3
LOCUS	AL517454/c
DEFINITION	AL517454 LTL_NF0L011_NB01 Homo sapiens CDNA clone CS0DA003Yf1l 3
ACCESSION	AL517454
VERSION	AL517454
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE	1 (bases 1 to 980)
JOURNAL	L.I.W.B., Gruber,C., Jesse,J. and Polayes,D.
COMMENT	Full-length cDNA libraries and normalization Unpublished (2001) Contact: Genoscope Genoscope - Centre National de Sequencage Bp 101 91006 Evry cedex - France Email: seqref@genoscope.cns.fr, web : www.genoscope.cns.fr. Location/Qualifiers 1..980
FEATURES	
SOURCE	/organism="Homo sapiens"

BASE COUNT  
ORIGIN

287 a 210 c 273 g 206 t 4 others

```
/db_xref="taxon:9606"  
/clone="CS0DA003FE11"  
/clone_lib="LTI-NFL011_NBC1"  
/sex="male"  
/tissue_type="neuroblastoma cells"  
/lab_host="DH10B"  
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA  
was primed with a NotI-oligo(dT) primer. Five prime end  
enriched, double-stranded cDNA was digested with Not I and  
cloned into the Not I and Eco RV sites of the pCMVSPORT 6  
vector. Library is not normalized, but is the control for  
the normalized libraries. Library was constructed by Life  
Technologies. Contact : Peng Liang Life Technologies, a  
division of Invitrogen 9800 Medical Center Drive Rockville  
Maryland 20850. USA Fax: (1) 301 610 8371 Email :  
fliang@lifetech.com URL :  
http://fulllength.invitrogen.com"
```

Query Match	24.9%;	Score 964.4;	DB 9;	Length 980;
Best Local Similarity	99.4%;	Pred. NO. 2.6e-137;		
Matches 974;	Conservative 4;	Mismatches 1;	Indels 1;	Gaps 1;

[illegible]

```
QY 1665 ggtttgcatgttggccaggctgtctcgaccttgcattgcagatgatccatgct 1724
DB 260 Ggtttgcatgttggccaggctgtctcgaccttgcattgcagatgatccatgct 201
QY 1725 tggctcccaagctgtgagttacagcgctgagccacacagccgctcttctaac 1784
DB 200 TGGCTCCCAAGCTGTGAGTBAAGCGCTGAGCCGACATGCCGCGCTTTCTCACC 141
QY 1785 tttaacactgtcttcttactcctcaatctgttttcaacacttcaatccctgtcttct 1844
DB 140 TTTACACTGTCTTCTTATCCTCAATCTGTTTACACACTTCAATCCCTGCTTCCAT 81
QY 1845 gttcaacactgtctcccaagctcatagctgctcttcttccatcttggtttaagggc 1904
DB 80 GTTACACTGTCTTCTTCCCAAGCTTCAATGCTGCTTCTTCAATTTGTTGAAGGCG 21
QY 1905 agtctctctgtgctgttt 1924
DB 20 AGTCTTCTCTGCTTGT 1

RESULT 4
AL571267/c 1031 bp mRNA linear EST 16-FEB-2001
LOCUS AL571267 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0D1015YK22 3
DEFINITION AL571267 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0D1015YK22 3
ACCESSION AL571267
VERSION AL571267.1 GI:12928392
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1031)
AUTHORS Li, W. B., Gruber, C., Jessee, J., and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
Source
1..1031
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0D1015YK22"
/clone_1lb="LTI_NFL006_PL2"
/tissue_type="placenta"
/note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@life.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 309 a 220 c 279 g 217 t 6 others
ORIGIN
```

```
Query Match 24.78; Score 955.4; DB 9; Length 1031;
Best Local Similarity 98.7%; Pred. No. 6e-136;
Matches 1020; Conservative 5; Mismatches 2; Indels 6; Gaps 6;
```

```
QY 1026 ctgctctttgataagaatgaagaagtcacagaacccgtggtatattaccagtaccac 1085
DB 913 CTGCTCTTTGATCAGAAAGACAGGTCAATGACAGAGAAACCGGTGATATTACAGTACAC 854
QY 1086 aggaacacttgggcccctctggaatgagacacttccagcccaacttgaagcagttc 1145
DB 853 AGGACACTTGGGGCCCTCTCGGATGAGCAGCTTCCAGCCCAACCTGAAAGCGTTCAAGTTC 794
QY 1146 tttaatgacacacttccagaagagctgtatgacttgtcttaaggaataataatgagt 1205
DB 793 TTTAATGATGACACACACTCCAGAAAGAGCTGATGACTTGTCTAAGGAATAATATGAT 734
QY 1206 gatgatgaggaagaagtgtggaataatgagacacttcttgaacttgaacgtggaaggaaccagc 1265
DB 733 GATGATGAGCGAGAAAGTTGGAATATGATGATGATGATGATGATGATGATGATGATGATGAT 674
QY 1266 tagcccaacagaacaaagaagacttcccttgggagcttgaagaaacagaatcttggc 1325
DB 673 TAGCCCAAGACCAAGAGAGACTTCCCTTGGCGTTGAGAAACAGATCTTGTGTC 614
QY 1326 ttltcccaacagccagctgttgataccctcagagccttcttacttccaaagtga 1385
DB 613 TTTTCCCAACAGCCAGGCTGTGATACCTGAGAGCCTTCTTACTCTCCAAAGTGA 554
QY 1386 aggaagccccgtctctcttaactgcatgcatcaagaagggtgagccttctctatctt 1445
DB 553 AGGGAAGCCCCGCTCTCTTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 494
QY 1446 cacacccgtccacactcagttacacactatcttctaac -tttttttgaatgagttc 1504
DB 493 CACACCTGCCCACCTCATGTTTACACCTCATGTTTACCTTTTACCTTTTGAATGAGT 434
QY 1505 cgcctctctgcccagcgtgga -gtgcaatgacagcttctcagcctcactgcaacctcgc 1563
DB 433 CGCTCTCTGCGCCAGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 374
QY 1564 tcttgggttcaagaactctgctgcatcagcctcccgagtaactctggaatgaagcagt 1623
DB 373 TCTTGGGTTCAGAAATCTCTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 314
QY 1624 gccacccagccagcgtatatttgaattttagtagaagaggttttggcatttggcca 1683
DB 313 GCCACCAAGCCCGGCTAATTTTGTATTTTGTATGATGATGATGATGATGATGATGATGATGAT 254
QY 1684 ggcgtgctcgaactcttgaactcagatgatactcctgacctggtgcccacagctgctg 1743
DB 253 GCGTGTGCTGCAACTCTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 194
QY 1744 gattacagcgctgagcacaatgcccgcgcctcttctcaacttcaacactgctctat 1803
DB 193 GATGACAGCGGTGAGCCACATGCGCGCGCTTTTTCACCTTTTACACTGCTCTTAT 134
QY 1804 cctcaactgttttcaacacttcaatccctgctctctc -atgttaacactgtctctcc 1862
DB 133 CCTACACTGTGTTTACACCTTATATCCCTGCTCTCTTATGTTTACACTGCTCTTCC 74
QY 1863 catgttaacagctccttcttcaacttggtttgaagggcagctctctctgctgt 1922
DB 73 CATGTTCATAGCTCTCTTCTTACATTTGGTTGAAGGCACTCTTCTGCTTGT 14
QY 1923 ttttggtttttc 1935
DB 13 TTTTGTGTTTCC 1

RESULT 5
AL546910 964 bp mRNA linear EST 16-FEB-2001
LOCUS AL546910 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0D1032YH24 5
DEFINITION AL546910 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0D1032YH24 5
ACCESSION AL546910
VERSION AL546910.1 GI:12880487
```





[illegible][illegible]



[illegible]

REFERENCE	(bases 1 to 925)
AUTHORS	Li,W.B., Gruber,C., Jesse,J. and Polayes,D.
TITLE	Full-length cDNA libraries and normalization
JOURNAL	unpublished (2001)
COMMENT	Contact : Genoscope Genoscope - Centre National de Sequencage BP 191 91006 Evry cedex - France Email: sequef@genoscope.cns.fr, web : www.genoscope.cns.fr.
FEATURES	Location/Qualifiers
source	1..925
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/clone="CS0D1065YP05"
	/clone_id="LTL_NFL006_PL2"
	/tissue_type="Placenta"
	/note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetechn.com URL : http://fulllength.invitrogen.com"
BASE COUNT	275 a 198 c 251 g 193 t 8 others
ORIGIN	
Query Match	23.7%; Score 917.4; DB 9; Length 925;
Best Local Similarity	98.9%; Pred. No. 3.7e-139;
Matches 915; Conservative	7; Mismatches 3; Indels 0; Gaps 0;
OY 1029	ctcttgatcagaatgacaaggtaagtcatagcgaagaaacctgggtattaccagracacag 1088
Dd 925	ccttttgatcagaatgacagagctcatgcagcagaaactgggtattaccagracacag 866
OY 1089	gacacttggggcccttcgatgatgacacttcccagcccaaacactgaaacagttaactt 1148
Dd 865	gaccttggggcccttcggatgatgacacttccagcccaaacactgaaacagttaactt 806
OY 1149	aatgtgaccaacatccagaagaagcgtgtatgacttgcctaagaataataatgatgat 1208
Dd 805	aatgtgaccacactctcacagagagcgctgatgacttgcctaagaataataatgatgat 746
OY 1209	gatggagggaagatgttgtgaatatgtgatagcccccttggaaacttggaggaagcagctag 1268
Dd 745	gatggagggaagatgttgtgaatatgtgtgacacttcttggaaacttggaggaagcagctag 686
OY 1269	cccaacagaacaaaagaagacttccctcttgcgttcacgaagaacacagatcttcttc 1328
Dd 685	cccaacagaacaaaagaagacttccctcttgcgttcacgaagaacacagatcttcttc 626
OY 1329	tcccaacagccccagcgtgtlgatalaccatcagaagccttctcttaactctccaagyaaag 1388
Dd 625	tcaccaacagccccagcgtgttgatgattaccctcagaagccttctcttaactctccaagyaaag 566
OY 1389	gaagccccgcgtctctctaactgcatgtatcataaggggttagcctgcttctctatctaac 1448
Dd 565	gaagccccgcgtctctctaactgcatgtatcatacaggggttagcctgcttctctatctaac 506
OY 1449	acctgcacacctatgttacaacctatcttccacaccttlttttgaatgtagtctgct 1508
Dd 505	acctgcacacctatgttacaacctatcttccacaccttlttttgaatgtagtctgct 446
OY 1509	ctcttgcacagcgtgagtgcaatggaacggtctcagctcaactgcgaacctccgcctctg 1568
Dd 445	ctcttgcacagcgtgagtgcaatggaacggtctcagctcaactgcgaacctccgcctctg 386
OY 1569	ggttcaaaccaatttcgtcgtcatcagccccccgagactctgagattaacaggaatgccaac 1628
Dd 385	ggttcaaaccaatttcgtcgtcatcagccccccgagactctgagattaacaggaatgccaac 326



Oy 1629 caagccggcgaattttagtatttagtagaagagggttttgcatgttgccagctg 1688  
Db 325 CACGCCCGGCTAATTTTGTATTTAGTAGAAGACGGGGTTTGGCATGTTGGCCAGGCTG 266  
Oy 1669 gtctgaactcttgacttcagatgagatcatctgcttgcccccacagtgtaggatta 1748  
Db 265 GTCTGAACCTTGTGACTTCAGATGATCCATCTGCTTGCCCTCCACAGTGTGGATGA 206  
Oy 1749 caggcgttagccacccatcccgccgctcttctacacttataccctgcttctacccca 1808  
Db 205 CAGCGKGKGGSSATGCGCGGCTCTTCTCACCCTTACACCGTGTCTTATATCCCA 146  
Oy 1809 catctgtttcaaccctcaatccctgctctcctcaatgttcaacattgcttcccatgt 1868  
Db 145 CATCTGTTTTCACACCTTCATCCCTGTCTCTCATGTTCACACTGTGTCTCCCATGTT 86  
Oy 1869 catagctccttcttaccatttggttgaaggcagctctctctgctgttttttg 1928  
Db 85 CATACCTGCTTCTTACCATTTTGTGTAAGGCGAGTCTCTGCTGTGTTTTTGTG 26  
Oy 1929 ttttccagaataatcagtatatt 1953  
Db 25 TTTTCCAGAAATCAGATATATT 1

RESULT 11  
BM471592 1053 bp mRNA linear EST 05-FEB-2002  
LOCUS AGENCOURT.646454 NIH\_MGC\_67 Homo sapiens cDNA clone IMAGE:5581397  
DEFINITION 5' mRNA sequence.  
ACCESSION BM471592.1 GI:18520634  
VERSION BM471592.1  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 1053)  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LMNL at:  
http://image.lnl.gov  
Plate: LHAM12341 row: d column: 06  
High quality sequence stop: 752.

## FEATURES

source  
1. 1053  
location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5581397"  
/clone\_id="NIH\_MGC\_67"  
/tissue\_type="retinoblastoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: eye; Vector: pCMV-SPORE6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.75 kb. Library constructed by Life  
Technologies."  
BASE COUNT 244 a 279 c 245 g 278 t 7 others  
ORIGIN

Query Match 23.2%; Score 896.6; DB 10; Length 1053;  
Best Local Similarity 96.9%; Pred. No. 5.3e-127;  
Matches 965; Conservative 0; Mismatches 24; Indels 7; Gaps 5;

Oy 747 atgagctgagccctccgacttcttaagaccttaagagtgcttcgacgctcgag 806

Db 1 ATGGAGCTGGCCCTCCGACTTCTTCAAAAGCTTTTACGATGCTCTGCGACCTGCAG 60  
Oy 807 gtctcaggagatcaagacttcaaggattctcacttccatagagatcatatgta 866  
Db 61 GGTTCAGGAGATATAGACTTCAAGATTCTTCACTTTCCATAGCATATATATGA 120  
Oy 867 gaagttctgaaatgcaaaatacagtgtagaagaaacccaacccaagttataggagctat 926  
Db 121 GAAGTTTCTGGAATGCAAAATACAGTGTGAAGAACCTACACCCAGTTATAGAGAGCTAT 180  
Oy 927 ccggtttggaatttgygttaccatgfatcatcttgcagtttgccatatataagtg 986  
Db 181 CCGGTTGAGAAATTTGTGGCTAACATGTATATTAATTGAGTTGGCTATATTAAGTTG 240  
Oy 987 aacgacctgaaatgtagcccccctgtagtcaagcctaaccgctcttgatcaaatgac 1046  
Db 241 AAGGACCTGGAAGATGACAGCCCTGCTGACGTACACTATCTGCTTTGATCAGAAATAC 300  
Oy 1047 aagttcatgcaagcaaaccttgtagtatatacagtaaccaaggacaattgggctctcg 1106  
Db 301 AAGGTATGACAGACAGAACTGGTATTAACAGTACACAGGAGACGTGGGCTCTCA 360  
Oy 1107 gatgagcacttccagcccaagacatgaagcagttcatttaagtatgacacactccag 1166  
Db 361 GATGAGCACTTCCAGCCAGACCTGAAACAGTCACTGTTCTTAATGTACACACACTCCAG 420  
Oy 1167 aagagcctgatatgaacttgcataaagaaataataatgatatgtagaaggaatgtg 1226  
Db 421 AAGGAGCTGTATGACTTTTGCTATAGAAATATATATGATGATGATGAGGAGAAATGTG 480  
Oy 1227 gaatatgtgatgacactcttggaactgtagagagaccagctagccacaagaacaagaag 1286  
Db 481 GAATATGTGATGATGACCTTTGGAACTGGAGGAGACACTGACCCACAGCAACAAAGAG 540  
Oy 1287 acttcctctggcgttcaaggaaacaatctctgtctcttcccaacacccagctg 1346  
Db 541 ACTTCCTTGGCGTTCAGGAACACAGATCTTGTCTTTCCACAGCCAGGCTG 600  
Oy 1347 ttgataactcaagagcctctcttactctcacaagtgaaggaagcccgctctcta 1406  
Db 601 TTGATACCTCAGAGCCTTCTCTTACTCTCCAAATGAAAGGAAGCCCGCTCTCTTA 660  
Oy 1407 actgcatgtcatcagggtgtagcctctctctactacttcaaccctgcacactatgttc 1466  
Db 661 ACTGATGTCAATCAGGGGTGAGCCCTTCCATCTTCAACACTGCACCTCATGTTC 720  
Oy 1467 aaactatcttctacactttttttagatgtagatctcgctctcttgcccaagctgag 1526  
Db 721 ACACCTATCTTCTCACCCTTTTGTGAGATGAGATCTCGCTCTTCCAGGCTGAG 780  
Oy 1527 tgcattgcaagttctcagctcaactcacc- gctctctgggttcaagcaattctgc 1585  
Db 781 TGCATGCGACGTTCTCAGCTCAGCTCAGCACTCCAGCCTCTTTGGGTCACGAAATCTGC 840  
Oy 1586 tgcatacagcctcccgagtaactg- ggaatacaggaatgtagccacaag- cccgagtaact 1643  
Db 841 TGCATCAGCCCTCCGAGTACCTGAGNATTTACAGGATGTGCACACAGCCCGGCTAANT 900  
Oy 1644 ttgtatttttagtagaga---cgagggttgcgaatgttgcccaagctgtgtctgaacct 1700  
Db 901 TTGATTTTAAANMANAAGACGGGGTTTGCATGTTTGGCAGAGCTGTCTCGAATCT 960  
Oy 1701 t-gaactcagatgacatcctgcttgccctccac 1735  
Db 961 TGGACTTCAGATGATCATCTGCTTGCTGCTCCCTCC 996

RESULT 12  
AL541167 916 bp mRNA linear EST 16-FEB-2001  
LOCUS AL541167 LTI\_FL002.P1 Homo sapiens cDNA clone CS0DE005F12 5 prime  
DEFINITION , mRNA sequence.







```

Db 420 GCGAGCCCTACAGTTCCTGAGTTCCGTTACTTCAAGGCAAAATACCTCCCAAGCA 479
QY 556 tcgcgcgtgtcacactttctactgaagcaatcctgatgcaaaatgatgaagaaaca 615
Db 480 TCGCGGTGTCACACCTTCTACTGAAAGCATCTGATGATGAATGATGAAGAGAAACA 539
QY 616 tggcatlataaagaactgctgtgtccgagagactcattaaagactggaaccagt 675
Db 540 TGGCATATATAGAKKCTGCTGTGTCGAGGACTCATTAAGACCTGGAACCAAGT 599
QY 676 cataagaacgcttltcatccgagcaagtgcgagcatacaacgctgagaaactggaacat 735
Db 600 CATATGAAGCCCTGTTCATCCGAGCAGTGGCGCATCAACAGTGAGACGAGAACAT 659
QY 736 ccatcaagaacatgagactgagccctcccgactcttcaaaagcctttaagatgtctcg 795
Db 660 CCATACAGACATGAGACTGGCCCTCCGACTTCTTCAAGCCCTTTACAGAGTCTCG 719
QY 796 cagctcgagaggttccagagagatacaagactcaagagattctacccttccataag 855
Db 720 CAGCTCGAGAGCTTCCAGGAGATCAAGACTTCAAGGATTTCTACTTTCCATAGCAG 779
QY 856 atcatalgtagaagttctggaatgcaaaatacagtgatgaagaagactcaccacagttta 915
Db 780 ATCATATGTAGAGTTCTGGAATGCAAAATACAGTGTGAAGAGAACCTCACCCAGTTA 839
QY 916 tagaagctatccggttggagaatttgggtaccatgatactactgactgactgct 975
Db 840 TAGG-KCTATCCGCTTGAG-AAATTTGGCTACCATGATACATACCTTGCAGTTTG-CT 896
QY 976 attaaagtga 987
Db 897 ATTATAGTTCA 908

```

```

RESULT 15
AL554750 956 bp MRNA linear EST 16-FEB-2001
LOCUS AL554750 LIT_NFL006_P12 Homo sapiens cDNA clone CS0DI086YJ04 5
DEFINITION AL554750 LIT_NFL006_P12 Homo sapiens cDNA clone CS0DI086YJ04 5
ACCESSION AL554750
VERSION AL554750.1 GI:12895833
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 956)
AUTHORS Li, W. B., Gruber, C., Jesse, J., and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
Source
1. 956
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DI086YJ04"
/clone_1db="LIT_NFL006_P12"
/tissue_type="placenta"
/note="Vector: PCWSPORT 6; Site_1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the PCWSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"

```

```

BASE COUNT 204 a 292 c 258 g 189 t 13 others
ORIGIN
Query Match 22.0%; Score 849.8; DB 9; Length 956;
Best Local Similarity 96.8%; Pred. No. 7.2e-120;
Matches 905; Conservative 13; Mismatches 10; Indels 7; Gaps 5;
QY 63 atggaagccggggcgccgggggggagcgctgctagcgctgctgctgctgctgctg 122
Db 1 ATGAGAGCCGGGGCCCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 60
QY 123 ctgcgcgcggggcgccccaatacgaacgtctaacg-ctccgagcttcccaaggagca 181
Db 61 CTGC-CGCGGGGGCC-CCCAATAGAACGCTACAGCGTTCGGCAGCTTCCACGGAGCA 118
QY 182 gctgagccgctcagtcagtcctacccggcagcgctggaagaagctagcagcgactg 241
Db 119 GCTGATGCCGCTCCAGTTCGAGTCCGCTACCGGACGCGCTGGACAAATACAGCGGAGCACTG 178
QY 242 g--gcgagagcstkkgctacttgagatacagctcgagcgtcgagcgcttgctcgag 299
Db 179 GCGCGGAGAGCGTGGGCTACTGAGATCAAGCTCGGGCTGACCGCTTCTGGCGAC 238
QY 300 agcgaagcctcttcgcaacgcgaactgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 357
Db 239 ACTAKKTCCTTCTGCAACCGCACTTCAKTCGCGGGCGGCAKCCCKAKCCCTGTGCGCG 298
QY 358 gctcgccagctatcccgagctgagcctctcgaggggcgtgctgagccgcgcgcgcgc 417
Db 299 GCTGCGCAGCTATCCGAGCTCGGCGCTCTGCGGGGGGGGGGGGGGGGGGGGGGGGG 358
QY 418 tcaagcgctggaagcaagcgctgcagcagcttcgcgcgcgcgcgcgcgcgcgcgcgcgc 477
Db 359 TCAAGCGCTGCAACGAGGCGCTGCGAGCTTCCGCACTTCCGCACTTCCGCACTTCCG 418
QY 478 tgcgcgacttccagcgccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 537
Db 419 TGGCGGACTTCCAGCGCGCGGAGCCCTACAGTTCTGCTACTTCAAGGAA 478
QY 538 ataactccccaagcaatcgcgctcgaacacttctactgaagaatcctgagcg 597
Db 479 ATATCTCCCAAGCAATCGCGCTGCTCAACACTTCTCAAGAGCATCTGATGATG 538
QY 598 aaatgataagaggaacatgcatatataaagacccgctgctgctgctgctgctgctg 657
Db 539 AAATGATGAGAGGAGCATGCGATATTAAGAGCTGCTGCTGCTGCTGCTGCTGCTGCT 598
QY 658 aagacctggaacccaagtcataatgaaagcctgttcacatcgagcagctgagcgacataca 717
Db 599 AAGACCTGGAACCAAGCATATGAAGCCGTTCATCCGAGCAGTGGGGGCAATACAAAG 658
QY 718 gtggaacttgagaacatcacaagacatgagctgagccctccgaacttctcaag 777
Db 659 GTGAGAACTGGGAACAATCCATCACAGACATGAGAGTGGCCCTTCCGACTTCTTCAAG 718
QY 778 cctttcagtgctctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 837
Db 719 CTTTTCAGAGTGTCTGCAAGCTGCGAGGGTTCAGGGAATPAAAGACTTCAAGGAT 778
QY 838 tctaccttccatagcagatcattatgtagaagttctggaatgcaaaatacagtgtag 897
Db 779 TCTACCTTTCATAGCAGATCATATATGAGAAGTTTTCGAATGCAAAATACAGTGTAG 838
QY 898 agaacctcaaccagttataggaagctatccggttggaatgtggtctacatgctatc 957
Db 839 AGAACCTCACCCAGTTATAGAGGCTATCGGTGGAATTTGTGCTACATGATC 898
QY 958 attactgagctgtgctctatataagtgaagcgc 992
Db 899 ATTACTTGCAGTTTGCCTATTATTAAGTTGAACGAC 933

```

Sat Jun 8 10:13:32 2002

us-09-729-674-1.rst

Page 14

Search completed: June 7, 2002, 22:07:46  
Job time: 4411 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 7, 2002, 20:58:20 ; Search time 6829.55 Seconds  
(without alignments)  
11861.195 Million cell updates/sec

Title: US-09-729-674-1  
Perfect score: 3871  
Sequence: 1 ttctctctctctctctctt.....ataaaaaaaaaaaaaa 3871

Scoring table: IDENTITY\_NUC  
Gapop 10.0 ; Gapext 1.0

Searched: 1797656 segs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_jm:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htg\_inv:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
------------	-------------	-------	--------	----	-------------

1	2237	57.8	2307	9	HSJ6470	AJ006470 Homo sapi
2	1955.4	50.5	1961	9	AK001634	AK001634 Homo sapi
3	1370.4	35.4	1439	9	BC008745	BC008745 Homo sapi
4	944	24.4	1685	10	MMAJ6469	AJ006469 Mus muscu
5	497.6	12.9	1665	5	GGCARR	X97607 G.gallus mR
6	467	12.1	1889	9	BC001047	BC001047 Homo sapi
7	467	12.1	2037	9	BC007942	BC007942 Homo sapi
8	467	12.1	2443	9	AK056085	AK056085 Homo sapi
9	466.6	12.1	1745	9	BC011701	BC011701 Homo sapi
10	465.8	12.0	2347	6	AX335490	AX335490 Sequence
11	465.8	12.0	2347	9	HSU47621	U47621 Homo sapien
12	465.4	12.0	2079	9	HSJ250583	AJ250583 Homo sapi
13	403	10.4	1407	10	RNSC65MR	X65454 R.norvegicu
14	369.4	9.5	407	11	G27784	G27784 human STS S
15	355.6	9.2	407	6	AX150162	AX150162 Sequence
16	313.6	8.1	330	6	AX150140	AX150140 Sequence
17	263.4	6.8	274	6	AX333823	AX333823 Sequence
18	258.6	6.7	279	6	AX261032	AX261032 Sequence
19	252.2	6.5	157326	9	AC008810	AC008810 Homo sapi
20	250.6	6.5	257967	2	AL365337	AL365337 Mus muscu
21	249.6	6.4	164217	9	AC093264	AC093264 Homo sapi
22	249	6.4	200240	2	AC011618	AC011618 Homo sapi
23	247.8	6.4	157375	9	AC025207	AC025207 Homo sapi
24	247.8	6.4	178420	9	AP003548	AP003548 Homo sapi
25	247.8	6.4	199569	9	AC016250	AC016250 Homo sapi
26	247.8	6.4	211076	2	AC012214	AC012214 Homo sapi
27	247.4	6.4	167330	2	AC021822	AC021822 Homo sapi
28	247.4	6.4	192361	2	AC025678	AC025678 Homo sapi
29	246.6	6.4	190615	2	AL390994	AL390994 Homo sapi
30	245.8	6.3	169089	9	AC008751	AC008751 Homo sapi
31	245.8	6.3	192441	2	AC073404	AC073404 Homo sapi
32	245.6	6.3	76141	2	AC024393	AC024393 Homo sapi
33	245.2	6.3	164758	2	AC104982	AC104982 Homo sapi
34	245.2	6.3	187086	2	AC087510	AC087510 Homo sapi
35	245.2	6.3	188659	2	AC022169	AC022169 Homo sapi
36	245	6.3	76141	2	AC024393	AC024393 Homo sapi
37	243.8	6.3	178328	9	AC092547	AC092547 Homo sapi
38	243.6	6.3	159597	9	AC008747	AC008747 Homo sapi
39	243.6	6.3	159738	2	AC058855	AC058855 Homo sapi
40	243.6	6.3	165841	2	AC011276	AC011276 Homo sapi
41	243.4	6.3	170805	9	AL133321	AL133321 Human DNA
42	243.4	6.3	196720	9	AP002026	AP002026 Homo sapi
43	242.8	6.3	145299	2	AF228728	AF228728 Homo sapi
44	242.8	6.3	169740	2	AF186193	AF186193 Homo sapi
45	242.8	6.3	175256	2	AC090775	AC090775 Homo sapi

#### ALIGNMENTS

RESULT 1  
LOCUS HSAJ6470 2307 bp mRNA linear PRI 23-MAR-2000  
DEFINITION Homo sapiens mRNA for cartilage-associated protein (CASP).  
ACCESSION AJ006470  
VERSION AJ006470.1 GI:3687321  
KEYWORDS cartilage-associated protein; CASP.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE  
AUTHORS Tonachini,L., Morello,R., Montlicone,M., Skaug,J., Scherer,S.W.,  
Cancedda,R. and Castagnola,P.  
TITLE cDNA cloning, characterization and chromosome mapping of the gene  
encoding human cartilage associated protein (CPRP)  
JOURNAL Cytogenet. Cell Genet. 87 (3-4), 191-194 (1999)  
MEDLINE 20169181  
REFERENCE  
AUTHORS Castagnola,P.  
TITLE Direct Submission  
JOURNAL Submitted (12-JUN-1998) Castagnola P., Biotechnologie in Oncologia,  
Ist. Naz.le per la Ricerca sul Cancro, L.go R. Benzi 10 Genoa,

16132, ITALY  
location/Qualifiers  
1. .2307  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="Contig of PCPH2 and I.M.A.G.E. cloneID 1090695"  
/issue\_type="fetal brain"  
/dev\_stage="19-23 weeks-old"  
12. .89  
/gene="casp"  
12. .1217  
/gene="casp"  
/codon\_start=-1  
/product="cartilage-associated protein (CASP)"  
/protein\_id="CAA07054.1"  
/db\_xref="GI:368732"  
/db\_xref="SPTREMBL:075718"  
/translation="MEPRGAAALLLIVACALRAGROYERSRSPDELMFL  
ESAYRHALDKYSGEHWAEVSYLEISLRRLRLDSEAFCHRNCSAAPPPEPAGLAS  
YPELRLEGGILLRRAHCLKRCQGLPAFROSPSREVLADQRRPEYSEFLFAFKAAN  
LPKIAAAHFTFLKHPDEMKRMAYKSLPGAEYIDKLETSYSEFLFAFKAAN  
GENMRSITDMELALPDFKAFYECLAAGSREIKDFKDYSLADHYEVEICKIO  
CEENLPVIGGYVEKPVATMYHYLOFAAYKINDIKRAACAVSYLLFDNDKMOON  
LYYQHRDWMGSLDEHFQRPPEAVQFFNVTTLOKELDYFAKENIMDDGEEVEYVD  
DLLELETS"  
12. .1217  
/gene="casp"  
90. .1214  
/gene="casp"  
/product="Cartilage-associated protein (CASP)"  
BASE COUNT 545 a 633 c 548 g 578 t 3 others  
ORIGIN

Query Match 57.8%; Score 2237; DB 9; Length 2307;  
Best Local Similarity 99.3%; Pred. No. 0;  
Matches 2263; Conservative 3; Mismatches 10; Indels 3; Gaps 2;

53 gccggcgagatgagccggggcgccggggcgagctgtctagcgtctgtgct 112  
|||  
2 GCACGAGGCGATGAGCGGGGCGCCGCGGGGCGCGCGCTGCTAGCGGTGCTGGCT 61  
Db  
113 ggcctgagcgtctgcgcgcggcgccgccaatacgaagctcgaagcttcgcaagcttc 172  
|||  
62 GGCCTGGCGCTGCGCGCGCGCGCCCAATACGAAGCTTCCACCTCCACCTTCC 121  
Db  
173 acggagacagctgtatgcgcgtcagctcagctaccgagcagcgcgtctgcaagtaacagcg 232  
|||  
122 ACGGAGAGAGCTGATGCCGCTCGAGTGGCTACCGGACGCGCTGAGCAAGTACAGCGG 181  
Db  
233 cgagcaactggcgagagacstlkgctacactlgaagatcaagcctgtgcgtcagcgttgc 292  
|||  
182 CGACACTGGGCGCGAGAGCGTGGGCTACTGGAATCAGCTCGGGCTGCACCGCTTCT 241  
Db  
293 ggcgagacagagagccttcctgcacacgaactgcagcgcgccgagcccgagccgcgc 352  
|||  
242 GCGCGAGAGCGAGGCTTTCACACCGCACTGAGCGCCGCGCGACGCCGAGCCGCGC 301  
Db  
353 cgcgcgctcgcagcgtatccgcagctgcgcgtcttcgaggcgctgtcgcgcgcgcgcgcgc 412  
|||  
302 CGCGGCGCTCGCCAGCTATCCGAGTGGCGCTTTCGCGGGCGCTGCTGCGCGCGCA 361  
Db  
413 ctgcctcaagcgtctgaagcagggcgtcagcgttcgcgcagctccagcagcccaagcgca 472  
|||  
362 CTGCTCAAGGCGTGAAGGAGGCGCTGCCAGCTTCCGCGCAGTCCAGCCAGCCGCGGA 421  
Db  
473 ggtgctggcgagcttccagcgccgagccctacaagttcctgcaagttcgttacttca 532  
|||  
422 GGTGCTGGCGACTTTCAGCGCCCGCGAGCCCTTACAGTTCCTGCAAGTTCGCTTCTCAA 481  
Db  
533 ggcacaatctcccaagacatcgcgctgtcacaaccttctactgaagcaacctcga 592  
|||  
482 GGCAAAATATCTCCCAAGGCAATCGCGCTGCTACACCTTTTACTGAGAGCACTCTGA 541  
Db

593 tgacgaatgtgaagagaaatgcatgtcatatataagagcctgtcgttgcgaagacta 652  
|||||  
542 TGACCAATGATGAGAGAGAACATGCAATATTATAAGAGCCCTGCTGCGGAGACTA 601  
Db  
653 catlaaagacctggaacaaagtcataatgaagcctgttcatccgagcagtgaggcata 712  
|||  
602 CATTAAGACCTGGAAACCAAGTCAATGAAGCCCTGTTCAATCCGAGAGTCCGCGCA 661  
Db  
713 caacggtgaagacttggaatacatcatcaagagacttgagccttgcgccttccgaattct 772  
|||||  
662 CAACGCTGAGAACTGGAGAACATCATCAGACATGAGAGCTGGCCCTTCCGACTTCT 721  
Db  
773 caaaccttttaagagtctcgcagcagcttggaaggttccagggagatcaagaaatc 832  
|||  
722 CAACACCTTTACAGAGTCTCGACAGCTGGGAGGCTTCAGAGGAGATCAAGGACTTCA 781  
Db  
833 ggaattctacacttccatagacagatcatatgtagaagttctgtgaatgcgaataacagtg 892  
|||  
782 GGATTCTACCTTCCATAGCAGATCATTAATGTAAGTTCGTAATCAAAATACAGTG 841  
Db  
893 tgaagagaaacctcaacccagttataagagctatccggttgaagaaattgtgtacagat 952  
|||||  
842 TGAAGAGAACTCAACCCAGTATAGAGAGCTATCCGTTGAGAAATTTGGCTACCAT 901  
Db  
953 gtaatcacttgcagcttgcctatataagttgaagcagccttgaagaaatgcagccccc 1012  
|||||  
902 GTATCATTAATGAGTGTGCTTATTAATTAATTAATTAATTAATTAATTAATTAATTA 961  
Db  
1013 tgcagtcagctatctgtccttcttgcagaaatgcagaaagctatgcagcagaaacctgtgta 1072  
|||||  
962 TGCAATCAGCTATCTGCTTGTGATCAAAATGACAAAGGTGATGACGCAAGAACCTGTGTA 1021  
Db  
1073 ttaccagtaaccacagggagacacttggggcctcgtgagatgagcacttccagcccaagctga 1132  
|||  
1022 TTACCAAGTACCAAGGAGCACTTGGGCGCTCTCGATAGACACTTCCAGCCAGACCTGA 1081  
Db  
1133 agcagttcagttctttaaigtgacacacccagaaagagctgtatgacttgcgtaagga 1192  
|||  
1082 AGCATTCAGTCTTTAATGACACACACTCAGAAAGAGCTGATGACTTGTGTAAGGA 1141  
Db  
1193 aaataataatgatatgagtgaggagaaagtgtgaaatgtgagaccccttggaaact 1252  
|||||  
1142 AAATATAATGATGATGATGAGGAGAAAGTGTGAAATATGATGATGATGATGATGAT 1201  
Db  
1253 ggaagagaccagctagcccaagcaacccaagagagacttccctctgtgcgttcaagaaacac 1312  
|||  
1202 GGAGGAGACAGCTAGCCCAAGCAACCAAGAGACTTCTCTGGGCTTCAAGGAACAC 1261  
Db  
1313 agattccttgccttcccaagcagccagcgtgtgataccctcaagaccccttctaac 1372  
|||||  
1262 AGATTCTTTGCTCTTTCCCAACAGCCAGGCTGTGATACCTCGAGACCTTCTCTTAC 1321  
Db  
1373 tctccaaagtgaaggaagcccccgtctctcttaactgcatgtacagaggggtgagcctg 1432  
|||  
1322 TCTCCAAAGTGAAGGAGAGCCCGCTCTCTAATCTCAATCTCAATCTCAATCTCAATCT 1381  
Db  
1433 ccttccatcttcaacacttccacactcatgttcaacacttcttccacacttcttct 1492  
|||  
1382 CTTTCCATCTTACACACTGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCACTG 1441  
Db  
1493 gagaatgagctgcctcctctgcagagcttgagtgcaatgagcaagcttccagctcactg 1552  
|||  
1442 GAGATGAGAGTCTGCTCTCTGCGCAGGCTGAGATGCAATGAGCAAGTCTCAGCTCACTG 1501  
Db  
1553 caactcgcctctgtgttgaagaacatctgcgtcaatcagcctcccgagtaactgtgat 1612  
|||  
1502 CAACCTCCGCTCTTGGTTGAACCAATTCGTGCAATCAACCTCCGAGATCACTGGGAT 1561  
Db  
1613 tacaggaatgtgcacacagcccgcaaatcttcatattta--gtagaagcggggttt 1670  
|||||  
1562 TACAGGATGTGCAACCAAGCCGCGCTAATTTTGTATTTTAAATGAAGAGCGGGTTT 1621  
Db

QY	1671	gcaatcttggcagaagcttgcgcgaaccttgaatcacaatgatcatatgcctt-99gc	1729
QY	1671	gcaatcttggcagaagcttgcgcgaaccttgaatcacaatgatcatatgcctt-99gc	1729
Db	1622	gccatgttggccagagcttgcgcgaaccttgaatcacaatgatcatatgcctt-99gc	1681
QY	1730	tcccacagtgctcggagattacaagcgcttgagcacaacatgcccggcctcttctcaccttac	1789
Db	1682	tcccacagtgctcggagattacaagcgcttgagcacaacatgcccggcctcttctcaccttac	1741
QY	1790	acctgtctcttatccctcacatctgttttcacaccttcaatccctgtcttctcctcatgttca	1849
Db	1742	acctgtctcttatccctcacatctgttttcacaccttcaatccctgtcttctcctcatgttca	1801
QY	1850	caatgtcttcccacatgctcacaatgctccttcttcaatcttcttgaaggagagctt	1909
Db	1802	caatgtcttcccacatgctcacaatgctccttcttcaatcttcttgaaggagagctt	1861
QY	1910	tctctgtctgtcttcttcttcttccacagaataacagtaattctttaaacaagaanaa	1969
Db	1862	tctctgtgtcttcttcttcttccacagaataacagtaattctttaaacaagaanaa	1921
QY	1970	catcttcagaagaatgataatgttgaaacatcccttctgtcttaattgtcttccagaattt	2029
Db	1922	catcttcagaagaatgataatgttgaaacatcccttctgtcttaattgtcttccagaattt	1981
QY	2030	agcttccttcttcccacatcccggaagaagtgtgaggaagacatagcttaattctccagcc	2089
Db	1982	agcttccttcttcccacatcccggaagaagtgtgaggaagacatagcttaattctccagcc	2041
QY	2090	tcaacatgtcttcaacttgcttgcttgcttgtaaccaattctagaccacacagaaacagat	2149
Db	2042	tcaacatgtcttcaacttgcttgcttgcttgtaaccaattctagaccacacagaaacagat	2101
QY	2150	tgaatgagagatgtagaatgtgcagaatgttgcttcttgccttcccaattgtcatcctcaaat	2209
Db	2102	tgaatgagagatgtagaatgtgcagaatgttgcttcttgccttcccaattgtcatcctcaaat	2161
QY	2210	acaagagttggccgagctcccatcttgaggacacatgcttaagtaagcttcccgagttggaaa	2269
Db	2162	acaagagttggccgagctcccatcttgaggacacatgcttaagtaagcttcccgagttggaaa	2221
QY	2270	aggaagaagccagagctgtctgactgacttaattcttcttccagtaataatttgaagta	2328
Db	2222	aggaagaagccagagctgtctgactgacttaattcttcttccagtaataatttgaagta	2280
RESULT 2			
AK001634			
LOCUS	AK001634	1961 bp	linear PRI 22-FEB-2000
DEFINITION	Homo sapiens cDNA FLJ10772 fls, clone NT2RP4000243, highly similar		
ACCESSION	AK001634		
VERSION	AK001634.1	GI:7023007	
KEYWORDS	oligo capping; fls (full insert sequence).		
SOURCE	Homo sapiens testicular carcinoma cell line:NT2 cDNA to mRNA, clone_11b:NT2RP4 clone:NT2RP4000243.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
AUTHORS	1 (siles)		
TITLE	Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y., Nishikawa,T., Nagai,K., Sugano,S., Aotsuka,S., Yoshikawa,Y., Matsunawa,H., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagahari,K., Masuno,Y. and Sasaki,N.		
JOURNAL	NEDO human cDNA sequencing project		
REFERENCE	Unpublished (2000)		
AUTHORS	2 (bases 1 to 1961)		
TITLE	Isogai,T. and Otsuki,T.		
JOURNAL	Direct Submission		
COMMENT	Submitted (16-FEB-2000) to the DDBJ/EMBL/Genbank databases. Takao Isogai, Heli Research Institute, Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3951, Fax:81-438-52-3952)		
	NEDO human cDNA sequencing project supported by Ministry of		

```

FEATURES
    source
        International Trade and Industry of Japan: cDNA full insert
        sequencing: Research Association for Biotechnology; cDNA library
        construction, 5'-6' 3'-end one pass sequencing and clone selection
        Helix Research Institute (supported by Japan Key Technology Center
        etc.) and Department of Virology, Institute of Medical Science,
        University of Tokyo.
        Location/Qualifiers
            1..1961
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /cell_line="NT2"
            /cell_type="teratocarcinoma"
            /clone="NT2RP4.000243"
            /clone_11b="NT2RP4"
            /note="cloning vector: PME18SFU3; mRNA from NT2 neuronal
            precursor cells after 2-weeks retinoic acid (RA)
            induction."
BASE COUNT
    417 a      579 c      481 g      484 t
ORIGIN

```

[illegible]

QY 744 gacatgagcctggccctcccgactctctcaaaagcctttacagatgctctgcagccgc 803  
 DB 721 GACATGAGCCTGGCCCTCCGACTCTTCAAAAGCCTTTTACGATGCTCTGCACCCGCG 780  
 QY 804 gaaggttccaaaggaatcaagacttcaagattctcaactctccatagcagatcatat 863  
 DB 781 GAGGTTCCAGGAGATCAAGGACTTCAAGGATTCTTACCTTTCCATAGCAGATCATAT 840  
 QY 864 gtagaagttcttggaatgcaaaaataacagtgtgaagagaacccctcaacccagttataagagc 923  
 DB 841 GTAGAAGTTCTGGAATGCAAAATACAGTGAAGAGAACCTCACCCAGTATATAGAGGC 900  
 QY 924 taccggttgaagaatttggctaccatgfatcatctacttggcaggttggcattataag 983  
 DB 901 TATCCGTTGGAATTTTGTGCTACATGATCATTTACTTGGCACTTTGCCATTTATAG 960  
 QY 984 ttgaagcacttgaagaatgagcagcccccgtgtcagtcagactatctgtctctttgataagat 1043  
 DB 961 TTGAACGACCTGGAAGATGCAAGCCCGCTGTGCAAGTCACTATCTGCTCTTTGATCAGAT 1020  
 QY 1044 gacaaagtcacatgacagcagaacctggtatattacagtaaccacagggagacattggggcctc 1103  
 DB 1021 GACAAAGTCATGACAGCAGAACCTGCTATTACAGTACCAAGAGGACACGTGGGCGCTC 1080  
 QY 1104 tcgagatgagcaactcccaagccagcctgaagcagttcagttcttaagtgaacacatc 1163  
 DB 1081 TCAGATGAGCACTTCACAGCCAGCCTGAGCAGTTCAGTTCATTATGATGACACACTC 1140  
 QY 1164 cagaagagcctgtatgacttctgaagaaataataatgatgatgatgaggaaggaagt 1223  
 DB 1141 CAGAAGAGCCTGTATGACTTTTGCTAAGAAATATATATGATGATGATGAGGAGACAGTT 1200  
 QY 1224 gtggaatatgtgatgacactcttggaacttggaagagagaccagtagccacagcaacaa 1283  
 DB 1201 GTGGAATATGTGGAGACCTCTTGGAACGTGAGGAGACAGTACCCACAGCAACCAA 1260  
 QY 1284 gagacttctcttgagcgtctcaagaaacagatcttgcctcttcccaaacagccag 1343  
 DB 1261 GAGACTTCTCTTGCGCTTCAGGAACACAGATCTTTGCTCTTTCCCAACAGCCCAAG 1320  
 QY 1344 ctgtgatcactcagaagcctctctcttactcctcaaaagtgaagggaaagcccgctctc 1403  
 DB 1321 CTGTGATACCTCAGAGCCTCTCTTACTCTCCAAAGTGAAGGAGGAGCCCGCTCTCT 1380  
 QY 1404 ctaactgatctcatcaaggggtgagccgctctctctactctcaactcgcacactatg 1463  
 DB 1381 CTAACGTGATGTCATCAGGGGAGACCGCTTCTCATTTTCACACCTGCCCCACTCTCATG 1440  
 QY 1464 ttcaacactatcttctcaacttttttttgaatgagatcgcgctctcttgcccgagctg 1523  
 DB 1441 TTCACACTATCTTCTCACCCTTTTGTGAGATGAGATCTCGCTCTTGCCCAAGGCTG 1500  
 QY 1524 gagtgcacatggcagcttctcagctcactgcaacctcgccctcttgggttcaagaactct 1583  
 DB 1501 GAGTGCATAGCAGCTTCTCAGCTCAGTCACGCAACCTCCGCTTGGTGAAGCAATCT 1560  
 QY 1584 gctgcacatcagcctcccgagatcctgggattacagggatggtccacacagcccgctaat 1643  
 DB 1561 GCTGCATCAGCCTCCCGAGTACCTGGGATACAGGCATGTGCCACACGCCCGGCTAAT 1620  
 QY 1644 ttgatattttagtagagcaggggttttgcaatgttgcccaagcaggtgtgtgaactctga 1703  
 DB 1621 TTGATATTTTATGATAGAGCGGGGTTTTCAGATGTTGGCCAGGCTGGTCTCGAATCTTGA 1680  
 QY 1704 cttaagatgacacatctccttgccctcccaagtgctgggattacagggctgaagccac 1763  
 DB 1681 CTTGAGATGATCANTCTCTTGGCTCTCCACAGTCTGGGATTAACAGCGTGAGACCAC 1740  
 QY 1764 atgcgcggcctcttctcaacttcaacactgtctcttatacctcaacatctgtttcaac 1823  
 DB 1741 ATGCGCGGCTCTTCTTCACTTACACCTTACACCTTCTTATCTTATCTGTTTTCACAC 1800  
 QY 1824 cttaacccctgtctctcccaagttcaacactgtctctcccaatgltcaacagtgccttct 1883

DB 1801 CTTCACTCCCTGCTTCTCAGTTCACACTGTCTTCCCATGTCATAGCTGCTTCT 1860  
 QY 1884 taccatttgggttgaagggcagctctctcgtgctgttttttttttcccaagaat 1943  
 DB 1861 TACATTTTGGTTGAAGGGAGCTCTCTCGCTGTGTTTTTTGTTTCCCAAGAAAT 1920  
 QY 1944 cagttatctttttaaataagaanaacatctcctgaagatg 1984  
 DB 1921 CAGTATTTATTTTAAATAGAAAAACATCTCTAGAGATG 1961  
 RESULT 3  
 LOCUS BC008745 1439 bp mRNA linear PRI 12-JUL-2001  
 DEFINITION Homo sapiens, cartilage associated protein, clone MGC:1926  
 IMAGE:3347384, mRNA, complete cds.  
 ACCESSION BC008745  
 VERSION BC008745.1 GI:14250580  
 KEYWORDS MGC.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 REFERENCE 1 (bases 1 to 1439)  
 AUTHORS Strausberg,R.  
 TITLE Direct Submission  
 JOURNAL Submitted (25-MAY-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  
 REMARK  
 COMMENT NIH-MGC Project URL: <http://mgc.ncl.nih.gov>  
 Contact: MGC help desk  
 Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
 Tissue Procurement: DCPD/DP  
 CDNA Library Preparation: Rubin Laboratory  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),  
 Gaithersburg, Maryland,  
 Web site: <http://www.nisc.nih.gov/>  
 Contact: [nisc\\_mgc@nih.gov](mailto:nisc_mgc@nih.gov)  
 Shevchenko,Y., Wetherby,K.D., Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S., Dietrich,N.L., Guan,X., Gupta,J., Ho,S.-T., Karlin,E., Legspt,R., Lin,M., Maduro,O.L., Mastello,C., Mastrian,S.D., McCloskey,J.C., McDowell,J., Pearson,R., Snyder,B., Stantiripop,S., Thomas,P.J., Tjongson,E.E., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A., Zhang,L.-H. and Green,E.D.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAL Plate: 5 Row: 1 Column: 24  
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 3687321.  
 FEATURES  
 source  
 location/Qualifiers  
 1. 1439  
 /organism="Homo sapiens"  
 /db\_xref="locusID:10491"  
 /db\_xref="taxon:9606"  
 /clone="MGC:1926 IMAGE:3347384"  
 /tissue\_type="Kidney, renal cell adenocarcinoma"  
 /clone\_id="NIH\_MGC\_14"  
 /lab\_host="DH10B-R"  
 /note="Vector: pOTB7"  
 36..1241  
 /codon\_start=1  
 /product="cartilage associated protein"  
 /protein\_id="AA08745.1"  
 /db\_xref="GI:14250581"  
 /translation="MEPGRGAALLLALCVACALRAGRAYERYSPSPDELMPL  
 EASVRLADKYSGEHMAESGYEISLRHLRLLDSAPFCRNCSAPOPAPAGLMS  
 YPELRALGGLLRHRLCKRCKGIPARROSPSDVLAQDRREPRYFLQPAFYRKAN





protein_id="CAA07053.1"	
/db_xref="GI:3687320"	
/db_xref="SPRTEMBL:O88698"	
/translation="MGPRSPAALLVLCGCAPTPGRCQGEYRFPFNPPDELMPLSAYNAHLDQYSGEHMAISVCIETPLSLHLHLRLHRLSEFCHRNCSAATPAPACPASHAELRFGSVLRRACDLCKKCGQLPAPFQSOSSVSLADPOOREPYKTLQFAFANLPKALAAAHYTLKHPDDEMKRNMEYKSLPGASDHIKIDLETKEYSELFVRAVAYNGEENWTSISDMELAPDFLKAIFYECLLACDESRLKDKDFYLSIADHYVLEQK INCEMTLTPVIGGVPEKFEVATMYHYLOFAVYKYLINDLKNAPCAVSLTLDQSDRWQMLVYYVYHRDKMGLSDHEHQPRLPEAVQFENVTTLQKELYDFAQHILMDDESEVEYVDDLLEFEESA"	
gene	72..1274
/gene="casp"	147..1271
mat_peptide	/gene="casp"
/product="cartilage-associated protein (CASP)"	
BASE COUNT	362 a 503 c 449 g 371 t
ORIGIN	

Query Match	24.4%	Score 944	DB 10	Length 1685
Best Local Similarity	83.6%	Pred. No. 1.2e-150		
Matches 1080	Conservative	2	Mismatches 207	Indels 3
				Gaps 1

[illegible]

Db	728	GAACAGGAAAGCTGCATTTCGACATGAGAGTCGGCGCTTCCGACTTCCTCAAGGCTTT	787
QY	782	ttacgaatgtctcgaagcctcgaggggttcacaggagatacaaggacttaaggatttcta	841
Db	788	CTACACAGTCCCTGGGTCCTCGAGGGGCTGGGGGATCAAGACATTCAAGACACTTCA	847
QY	842	ccttcacatagacagatcatatgtagaattctcggatgcaaatatacagtgtagaagaa	901
Db	848	CCTGTCCATAGACAGATCACTATGTGTGAAGTTCTGGAGTGAAGATTCTGTTCAGAGAC	907
QY	902	cctcaacccagttatagagagctatccggttgagaaattgltggtccacatgatacta	961
Db	908	CCTCACCCACAGTCATAGAGGGCTATCCCTGGAGAAATTGTGGCGACCATGTACCACTA	967
QY	962	cttgaatttgcctttataagtltcaagacttgaagaatgagccccctgtgacatgag	1021
Db	968	TTTAAAGTTTGCTATTACAAAGTTGAATGATGTGAAGAAATGCAAGCCCCGTCCTCA	1027
QY	1022	ctatctgctctttgtatcagaatgacaaggltcatgacagaaacctglttatcacagta	1081
Db	1028	CTACCTGCTTTGACACAGAGTGACAGGGGTATGCAACAGAACCTGTACTATATAGTA	1087
QY	1082	ccacagggagaaacttgygggcctctcgatgtagaaccttccagccccagagccggaagcagttca	1141
Db	1088	CCACGGGGCAAGTGGGGCTCTCGAGTAGACACTTCCAGCCACGACCCGGAAGCACTTTA	1147
QY	1142	gtctcttaatgtagacacactccagaagaagctgtatgactctgtctaaagaaataataat	1201
Db	1148	GTTCTTTATATGTAGACAGACCTCTCACAAGAACTGTACGACTTCGCTCAAGAACACTTAAT	1207
QY	1202	ggaatgacatgagggagaaagtgtggaataatgttgatgtagacctctcttggaaactggaagagac	1261
Db	1208	GGATGACGATGAGGGAGAGGTTGTGGAGATGTGGACGCACTTGTGGAGACGGAAGACATC	1267
QY	1262	cagctagacccaagcaaaccaagaagactccct	1293
Db	1268	TGCTTAGTCCACAGGGGCTTAAGGAACCTTCT	1299

```

RESULT 5
LOCUS      GGCARRP                               1665 bp      mRNA      linear      VRT 26-JUN-1997
DEFINITION G.gallus mRNA for cartilage associated protein.
ACCESSION  X97607
VERSION    X97607.1  GI:1296525
KEYWORDS   cartilage associated protein.
SOURCE     chicken.
ORGANISM   Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE  1 (bases 1 to 1665)
AUTHORS   Castagnola,P., Genari,M., Morello,R., Tonachini,L., Marin,O.,
            Gargero,A. and Canceda,R.
TITLE      Cartilage associated protein (CASP) is a novel developmentally
            regulated chick embryo protein
JOURNAL    J. Cell. Sci. 110 (Pt 12), 1351-1359 (1997)
MEDLINE    97360293
REFERENCE  2 (bases 1 to 1665)
AUTHORS   Castagnola,P.
TITLE      Direct Submission
JOURNAL    Submitted (29-APR-1996) P. Castagnola, Centro di Biotecnologie
            Avanzate, Differenziamento Cellulare, L.go R. Benzi, 10 Genova,
            16132, ITALY

FEATURES
    source          Location/Qualifiers
                    1..1665
                        /organism="Gallus gallus"
                        /db_xref="taxon:9031"
                        /cell_type="chondrocyte"
                        /dev_stage="embryo"
                        /lab_host="E.coli"
                    43..858

```

```

/codon_start=1
/product="cartilage associated protein"
/protein_id="CA65206.1"
/db_xref="GI:1295526"
/db_xref="SWISS-PROT:Q90830"
/translation="MWRLLAALATLGAQERYSPKSPDEIMPLESAARYGLDQY
STEMESVSAYLEVMRLYRLRLDTEFAHCNCSAGPLAPPADDELAELRLAV
LRRKQCLRRCKQGLPAPRQAQPGRELLLEERQREPYLQFAVKAANLPRKAIAAAT
FLIKHDPDEMMORNMAYASIDPAEEHIIKDLERPYNLPLRAVRAKNGNPSISD
MELALPFEFTYDDCIACAGCSREIKRKFYLIADHYLEVLACKQDITTT"
BASE COUNT      470 a      385 c      388 g      418 t
ORIGIN

```

[illegible]

Result	6	1889 bp	mRNA	Linear	PRI 12-JUL-2001
LOCUS	BC001047				
DEFINITION	BC001047 Homo sapiens, nucleolar autoantigen (55kd) similar to rat synaptonemal complex protein, clone MGC:1322 IMAGE:3510406, mRNA, complete cds.				
ACCESSION	BC001047				
VERSION	BC001047.1				
KEYWORDS	MGC.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 1889)				
JOURNAL	Strausberg, R				
TITLE	Direct Submission				
COMMENT	Submitted (11-DEC-2000) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA				
REMARK	NIH-MGC Project URL: <a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a>				
COMMENT	Contact: MGC help desk Email: <a href="mailto:gcgaps-remail.nih.gov">gcgaps-remail.nih.gov</a> Tissue Procurement: ATCC DNA Library Preparation: Rubin Laboratory CNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Institute for Systems Biology <a href="http://www.systemsbio.org">http://www.systemsbio.org</a> contact: <a href="mailto:amandansystemsbiology.org">amandansystemsbiology.org</a> Anup Madan, Rachel Dickhoff, Jessica Fahy, Stephanie Ford, Julia Greene, Mark Kettelman and Anuradha Madan				
FEATURES	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <a href="http://image.lnl.gov">http://image.lnl.gov</a> Series: IRAL Plate: 7 Row: f Column: 22 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 5454037.				
SOURCE	Location/Qualifiers 1. 1889 /organism="Homo sapiens" /db_xref="LOCUSID:10609" /db_xref="taxon:9606" /clone="MGC:1322 IMAGE:3510406" /tissue_type="Brain, neuroblastoma" /clone_id="NIH_MGC_19" /lab_host="DH10B-R" /note="Vector: pOTB7" 47. 1360 /codon_start=1 /product="nucleolar autoantigen (55kd) similar to rat synaptonemal complex protein" /protein_id="AAH01047.1" /db_xref="GI:12654443" /translation="MARVAMGLIMLLGSGAOTERYKSPFGPPEDIMPLAAAYGAAL EYEGESRESANYLEALRLHLRLRSEAFCHANSQPAAPKPPDGGRADEMCCE LRLFGRLERACLRCKRTLPAPQVYPPROLRFQSRPLPYQYLHAYLFKRNREK AVAAAYFLIDRNKRHELTAKYLYNGQMLVDSESLDLEAPQYEAVFLFAVLYKNGS DFNSSTEDMERALSEYLAIVAPARCAGEAGHEQVDPEDFPATADLEASLQCKVGE ANLIPNNGYGVFNDKPYVATWYHYLOFAPAYKIKNDYROARSASAYMLDPKDSWQONVY YRFHRRNRGLSEEDPQPREANLTHNOTRELKELLEFTMTYQSDDEKLEETPEPL EPDADLSDAFFEEBEGDEBEGVYDVWQDEPAKDEAEAPEPELA"				
CDS					
BASE COUNT	399 a	570 c	611 g	309 t	
ORIGIN					
Query Match	12.1%	Score 467	DB 9	Length 1889	
Best Local Similarity	64.1%	Pred. NO. 9.7e-70			

[illegible]

Oy	1132	aagcagtcctagtcttctaattgtagccacactccagaagaagctgtatgccttgtaagg	1191
Dd	1110	AAGCCATGCTGTACCAACAACAGCAGCGCGACTGGGGAAGTGCCTGGATTGCCACCACA	1169
Oy	1192	aaataatagatgatgtaggaaggaagcttgtga	1228
Dd	1170	TGTACTGCAGTCACATCATGATGAGATGGAGACTTGAGAGA	1206
RESULT	7		
BC007942		2037 bp mRNA linear PRI 12-JUL-2001	
LOCUS	Homo sapiens, nucleolar autoantigen (55kD) similar to rat synaptosomal complex protein, clone MGC:14267 IMAGE:4130726, mRNA, complete cds.		
DEFINITION	BC007942		
ACCESSION	BC007942		
VERSION	BC007942.1 GI:14044027		
KEYWORDS	MGC.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
JOURNAL	1 (bases 1 to 2037) Straussberg,R. Direct Submission Submitted (11-MAY-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA		
REMARK	NIH-MGC Project URL: http://mgc.ncl.nih.gov		
COMMENT	Contact: MGC help desk Email: cgabs-rehall.nih.gov Tissue Procurement: ATCC CDNA Library Preparation: Rubin Laboratory CDNA library Arrayed by: The I.M.A.G.E. Consortium (LNLU) DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland: Web site: http://www.nisc.nih.gov/ Contact: nisc-mgc@ngri.nih.gov Shewchenko,Y., Wetherby,K.D., Beckstrom-Sternberg,S'M., Benjamin,B., Blakesley,R.W., Bouffard,G.C., Brinkley,C., Brooks,S., Dietrich,N.L., Guan,X., Gupta,J., Ho,S.-L., Karlins,E., Legaspi,R., Lim,M., Maduro,Q.L., Masello,C., Mastrian,S.D., McCloskey,J.C., McDowell,J., Pearson,R., Snyder,B., Stantiripop,S., Thomas,P.J., Tongsonn,E.E., Touchman,J.W., Tsurgeon,C., Vogl,T.L., Walker,M.A., Zhang,L.H. and Green,E.D.		
FEATURES	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLU at: http://lmge.lnl.gov Series: IRAL Plate: 20 Row: h Column: 16 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, Similarity but not identity to protein.		
SOURCE	location/Qualifiers		
	1..2037		
	/organism="Homo sapiens"		
	/db_xref="locusID:10609"		
	/db_xref="taxon:9606"		
	/clone="MGC:14267 IMAGE:4130726"		
	/tissue_type="Brain, neuroblastoma"		
	/clone_id="NH_MGC_19"		
	/lab_host="DHIOB-R"		
	/note="vector: pORB7"		
	42..1355		
	/codon_start=1		
	/product="nucleolar autoantigen (55kd) similar to rat synaptosomal complex protein"		
	/protein_id="AAH07942.1"		
	/db_xref="GI:14044028"		
	/translation="MARAWGLMLLLGSAGAOYEKSFRCPPEDIMPLPAAVGHALLBOYGEWSRESARYLEALARHLRLRILROSEAFCHANCSPAPPAKPDDPGRADWACE		
CDS			















Dh 669 CCCACAGGACCTAGAGCCCAAGCCCTTACGAGCCGCTTCTCCGCGCTTGAAGCTCT 728  
Qy 712 acaacggtgagaaactgagaaacatccatcacagacatgagagctggccctccgactct 771  
Dh 729 ACAACAGGGGAGATTCCGACGACGACGAGACATGAGACGGGCGCTTGCAGAGTACC 788  
Qy 772 tcaaacgctttaaagatgctgcagcctgcgagaggttccagaggagatcaagacttca 831  
Dh 789 TGGAGCTTTGGCCCGGCTGGCCGCTGTGAAGGGGCGCCATGAGCAGGTGGCTTCA 848  
Qy 832 aggaattctaccttccatagacatcatatgtagaaggtctgtaacgaaataacagt 891  
Dh 849 AGGACTTACCCGCGCATACACATCTCTTTCAGAGACTCCCTGCAGTGCAGGTGACT 908  
Qy 892 gtgaagaagacccacccagctatagagagctaccggttgagaaattgttgctacca 951  
Dh 909 GTGAGGCCAATTTGACCCCAATGTGGGTGCTACTTGTGAGACAAGTCTGGCGACCA 968  
Qy 952 tgtatcatctgacgttgccttataatgtagaagacccgtaagaatgcaagccct 1011  
Dh 969 TGTACACTACTCTGAGTTTCCCTACTATAAGTTGAAGATGTGCGCCAGGCTGCCCA 1028  
Qy 1012 gtgcagacgacatctgctcttctgacgaatgacaaggtcatgcaagacactgtgt 1071  
Dh 1029 GCGCGCCAGCTACATGCTCTTTCAGACCCCAAGACAGCGTATGACAGAACTGGTGT 1088  
Qy 1072 attacgaagacagagagacactgtgagcctcgcgagtagaaccttccagccagact 1131  
Dh 1089 ATTAGCCGTTTCCACGCGGCTGCTGCGGCTGGAAGAGAGAGAGACTTCCAGCGCGGAG 1148  
Qy 1132 aagcagtgcttctttaaagtgcacacacccaggaagagctgtagacttgcataag 1191  
Dh 1149 AGGCCATGCTTACACACACACGCGCGAGCTGCGGAGACTCTGAGTTTACCCACA 1208  
Qy 1192 aaatataatgagatgagtagaggaagagtggtgga 1228  
Dh 1209 TGTACTGCTGACATGATGATGATGATGAGCTGGAGGA 1245

RESULT 13  
RNSC65MR 1407 bp mRNA linear ROD 09-AUG-1996  
LOCUS R.norvegicus mRNA for SC65 synaptonemal complex protein.  
ACCESSION X65454  
VERSION X65454.1 GI:57191  
KEYWORDS synaptonemal complex.  
SOURCE Norway rat.  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
1 (bases 1 to 1407)  
Pearlman, R.E.  
REFERENCE  
AUTHORS Direct Submission  
TITLE Submitted (06-APR-1992) R.E. Pearlman, York University, Dept. of  
JOURNAL Biology, Farquharson Bldg., 4700 Keele Str., Downsview, Ontario,  
M3J 1P3, CANADA  
2 (bases 1 to 1407)  
Chen, Q., Pearlman, R.E. and Moens, P.B.  
REFERENCE Isolation and characterization of a cDNA encoding a synaptonemal  
AUTHORS complex protein  
TITLE  
JOURNAL Blochem. Cell Biol. 70 (10-11), 1030-1038 (1992)  
MEDLINE 93213429  
FEATURES  
source  
1..1407  
Location/Qualifiers  
/organism="Rattus norvegicus"  
/db\_xref="taxon:10116"  
/tissue\_type="testis"  
20..1315  
/codon\_start=1  
/product="SC65 synaptonemal complex protein"  
/protein\_id="CAA46449.1"  
/db\_xref="GI:57192"

Qy 172 cagcggagacgtagtgcgcgtgcagtgagctccgcgcgcgcgtgagacagacagc 231  
Dh 6 CCGCGAGAGACTGATGCTCTGGCCACGCGCTACGGCCAGCTCTGAGCAGATGAG 65  
Qy 232 ggcagacatggcgagagacatkgctacatcagatcagcctgcgcgtgcacgcctg 291  
Dh 66 GGAAGAGCTGGCGGAGAGCGCGGCTACTGAGAGCGCGCGCTGCGCTGATGGCTGC 125  
Qy 292 tgcgcgacagcagagccttctgcacgcgcacatgcag-----gcgcgcgcgcgcgc 345  
Dh 126 TGCGCGAGATGAGAGCGCTTTCGACGCGCACTGACGCGCGCGCGCGCGCGCGCG 185  
Qy 346 agccgc 378  
Dh 186 GCGCGCGCTTCCG 245  
Qy 379 tgcgccttctgc 438  
Dh 246 TGCGGCTTTCG 305  
Qy 439 tgcgcgccttccgc 498  
Dh 306 TGCGCGCGCTTCCG 365  
Qy 499 agccctcagatctctgcagtgcttcttcttcttcttcttcttcttcttcttcttct 558  
Dh 366 TGCGCGCGCTTCCG 425  
Qy 559 ccgcgtcgcacaccttcttcttcttcttcttcttcttcttcttcttcttcttcttct 618  
Dh 426 CTGCAGCTTACCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 485  
Qy 619 cataataagagcctgcgtgt---gcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 675  
Dh 486 ACTACTACAGGAGGAGTCTGATATGAGATGAGTCTTCTTCTTCTTCTTCTTCTTCT 545  
Qy 676 catatgaagcctgttcttcttcttcttcttcttcttcttcttcttcttcttcttct 735  
Dh 546 CTTACAGGCGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 605  
Qy 736 ccacacagacatgagctgcgccttccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 795  
Dh 606 GCACAGAACATGAGAGCGCGCTTGGCTGATTTATGACCGCTTTTGTCTGGGTCTAG 665  
Qy 796 cagcctgcagaggttcagagagatcaagagctcaagagcttcttcttcttcttcttct 855  
Dh 666 CTGCGCTTGAAGAGGCG 725  
Qy 856 atcattatgtagaagcttcttcttcttcttcttcttcttcttcttcttcttcttct 915  
Dh 726 ATCTCTTCGAGCAATTCCTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 785  
Qy 916 taggagcctacgcgttgagaaattgttgcttaccatgatacttcttcttcttcttct 975  
Dh 786 TAGGTGCTTCTTGTGAGCAAGTTCGTGGCCACCAATGATACCTACTGACGAGTTTCT 845